AD-A033 429 STANFORD UNIV CALIF DEPT OF STATISTICS F/G 12/2
OPTIMIZATION OF RECURRENT STOCHASTIC CAPACITY EXPANSION MODELS --ETC(U)
OCT 76 R S SHIPLEY N00014-75-C-0561
UNCLASSIFIED TR-180 NL

1 OF 1
AD
A033429

END
DATE
FILMED
2-77

12

OPTIMIZATION OF RECURRENT STOCHASTIC CAPACITY
EXPANSION MODELS AND GENERALIZATION
TO A NON-RECURRENT MODEL

by

R. SCOTT SHIPLEY

TECHNICAL REPORT NO. 180
October 11, 1976

SUPPORTED BY THE ARMY AND NAVY
UNDER CONTRACT N00014-75-C-0561 (NR-042-002)
WITH THE OFFICE OF NAVAL RESEARCH

Gerald J. Lieberman, Project Director

DEPARTMENT OF OPERATIONS RESEARCH
AND
DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1. Summary of Previous Results

In two previous papers ([5] and [6]), stochastic capacity expansion Models I and II were introduced. Both models permit two types of facilities: permanent and temporary. Permanent facilities represent the means by which demand is normally accommodated, while temporary facilities represent the extraordinary measures taken in order to accommodate excess demand prior to an expansion of permanent facilities. Examples of temporary facility usage include cases of backlogging, overloading and jobletting.

Models I and II were constructed under the following assumptions:

a. The demand process and costs are time-stationary.

b. The demand process can be characterized as the difference of two simple Poisson processes (without truncation).*

c. All costs are independent of the amount of permanent capacity available.

d. Permanent facility operating costs are proportional to the amount of permanent facility capacity actually utilized.**

---

*The untruncated demand assumption continues to permit departure when the demand level is zero. The viability of this assumption is discussed at length in Section 1.7 of [5].

**It is demonstrated in Sections 1.5 of both [5] and [6] that permanent facility operating cost terms which are proportional to the permanent capacity available are also permissible.

Model I, introduced in [5], addresses the case where temporary facilities are non-modular with usage costs depending solely on the level of excess demand being served. Model II, introduced in [6], treats the case of modular temporary facilities which are typically available in a fixed increment size (a module) and incur instantaneous installation and removal charges in addition to normal usage costs. The objective in each model is the minimization of expected total discounted costs. The expansion decision strategies considered are of the recurrent form (X,K), where K denotes a limit on temporary facility usage and $X+1$ denotes the permanent facility expansion size. That is, whenever excess demand equals K, a permanent expansion of size $X+1$ is implemented upon the arrival of an additional demand unit.

In each model, the assumptions (a) - (d) allowed the derivation of a single finite linear system of expected cost equations for fixed values of X and K. In the case of Model II, the derivation of the linear equation system was preceeded by a demonstration that an optimal module-removal policy is of the form: "remove a module only if at least s^* units of unused capacity will remain available". An algorithm was given to determine the optimal parameter s^*.

By parameterizing the expected cost equation systems in the expansion variable X, it was shown that linear equation solution techniques can be used to derive expected cost functionals $C_0(\cdot,K)$ for each fixed value K. It was also shown that an optimal expansion size $X^*(K) + 1$ satisfies

$$C_0(X^*(K), K) = \min\{C_0(X,K) : \text{integer } X \geq K\} .$$

Algorithms were given in [5] and [6] for recursively computing these functionals over all feasible values for K. The form of the functionals was identical for both Models I and II.

This paper consists of two self-contained parts. In Chapter 2, the determination of optimal expansion sizes for recurrent Models I and II is considered. In Chapter 3, a revised non-recurrent model, permitting the relaxation of assumptions (b) - (d) above, is introduced.

1.2. Notation

$\mathcal{D}(t)$ will denote the demand at time $t \geq 0$. Given Poisson arrival rate $\lambda_1 > 0$ and Poisson departure rate $\lambda_2 > 0$, $\lambda = \lambda_1 + \lambda_2$ will denote the combined Poisson event rate, with arrival probability $p = \lambda_1/\lambda > 0$ and departure probability $q = \lambda_2/\lambda > 0 (p+q = 1)$ [3]. is, whenever an event occurs, the event is a unit demand increase with probability p and the event is a unit demand decrease with probability q.

As introduced earlier in this chapter, K will benote the temporary facilities usage limit and $X+1$ will denote the permanent facilities expansion size. g will denote the expansion cost function, where $g(X)$ represents the cost of an expansion of size $X+1$. r will denote the continuous discount rate; $0 < r < 1$.

k will denote the spares level: the permanent capacity available less demand. Thus, whenever $k \geq 0$, the permanent facilities available

suffice to accommodate demand. However, whenever $k < 0$, $-k$ represents excess demand that must be accommodated by temporary facilities.

$\mathbb{R}^m$ will denote the space of all real vectors having m components, $m \geq 1$. A vector is a column vector unless otherwise noted. Given $V \in \mathbb{R}^m$, V^T will denote the transposed vector. $\mathbb{R}^{m \times n}$ will denote the space of all real matrices having m rows and n columns. Given $A \in \mathbb{R}^{m \times n}$, $A_{i\cdot}$ will denote row i of A and $A_{\cdot j}$ will similarly denote column j of A ($A_{i\cdot}^T \in \mathbb{R}^n$ and $A_{\cdot j} \in \mathbb{R}^m$).

CHAPTER 2

COMPUTING OPTIMAL EXPANSION SIZES FOR THE RECURRENT MODELS

2.1. Introduction

[5] and [6] demonstrate that for a fixed temporary facility usage limit $K \in \kappa$ (where κ denotes a nonnegative integer set of feasible limits with $0 \in \kappa$), an optimal permanent expansion size $X^*(K) + 1$ satisfies

$$C_0(X^*(K), K) = \min\{C_0(X,K) : \text{integer } X \geq K\} \tag{2.0}$$

where

$$C_0(X,K) = \frac{\emptyset(K) + g(X)}{Z^{-1} - Z^X} Z^K, \qquad X \geq K; \tag{2.1}$$

$$\emptyset(\cdot) \text{ is nonnegative and nondecreasing over } \kappa;^* \tag{2.2}$$

$$0 < Z = \frac{1 - \sqrt{1 - 4\alpha\beta}}{2\alpha} < 1, \text{ given } \alpha = \lambda q(\lambda+r)^{-1}$$

$$\text{and } \beta = \lambda p(\lambda+r)^{-1} \quad (Z \text{ satisfies } \alpha Z^2 - Z + \beta = 0); \tag{2.3}$$

$C_k(X,K) = Z^k C_0(X,K)$, where $C_k(X,K)$ denotes the expected total discounted costs starting from initial spares level $k \geq 0$ (with no temporary facilities initially in use)

---

*Algorithms for computing $\{\emptyset(K) : K \in \kappa\}$ are given in [5] and [6].

and using permanent expansion size $X+1$ $(X \geq K)$, when the temporary facilities usage limit has value $K \in \kappa$; (2.4)

$Z^k = E[e^{-rT_k}]$, where T_k denotes the first time that the net demand increases by an amount $k \geq 1$. (2.5)

This chapter treats the above problem under two basic assumptions:

Assumption 2.1. g is nonnegative and nondecreasing over $[0,\infty)$.

Assumption 2.2. There exists a known, finite, nondecreasing function $U(\cdot)$ over κ such that $X^*(K) \in [K, U(K)]$, $K \in \kappa$.

Assumption 2.1 is self-explanatory and anything to the contrary would be unrealistic. Assumption 2.2 is used to insure that the search for $X^*(K)$ can be restricted to the finite interval $[K, U(K)]$. In practical situations, unlimited expansion is obviously infeasible. The following lemma provides mathematical justification for this point of view.

Lemma 2.1. If there exists $\Delta > 0$ such that $g(X+\Delta) - g(X) \geq \xi > 0$ for all $X \geq 0$, then Assumption 2.2 is satisfied with

$$U(K) = \max\{K,\ U'(K),\ [\frac{\Delta}{\xi}(\frac{Z^{-1}(\phi(K) + g(K))}{Z^{-1} - Z^K} - \phi(K)) + 1]\}, \quad K \in \kappa$$

where $U'(0) = 0$, and

$$U'(K) = \max\{U(K') : K' \in \kappa \sim [K,\infty)\}\ , \qquad K \in \kappa \sim \{0\}$$

(the brackets denote integer truncation).

Proof. The inclusion of $U'(\cdot)$ insures that $U(\cdot)$ is nondecreasing over κ. Suppose $X > U(K)$. Then

$$\frac{X}{\Delta} > \frac{1}{\xi}\left(\frac{Z^{-1}(\phi(K) + g(K))}{Z^{-1} - Z^{K}} - \phi(K)\right) + 1$$

$$\Rightarrow \left[\frac{X}{\Delta}\right] > \frac{1}{\xi}\left(\frac{Z^{-1}(\phi(K) + g(K))}{Z^{-1} - Z^{K}} - \phi(K)\right)$$

$$\Rightarrow g(X) \geq g(X) - g(0) \geq \xi\left[\frac{X}{\Delta}\right] > \frac{Z^{-1}(\phi(K) + g(K))}{Z^{-1} - Z^{K}} - \phi(K)$$

$$\Rightarrow \phi(K) + g(X) > \frac{(Z^{-1} - Z^{X})\,(\phi(K) + g(K))}{Z^{-1} - Z^{K}}$$

$$\Rightarrow \frac{\phi(K) + g(X)}{Z^{-1} - Z^{X}} > \frac{\phi(K) + g(K)}{Z^{-1} - Z^{K}}$$

$$\Rightarrow C_0(X,K) > C_0(K,K)\ . \qquad \square$$

The above lemma essentially states that unlimited expansion is never optimal if the marginal costs of expansion are bounded away from zero.

Denote

$$C(X,K) = \frac{\phi(K) + g(X)}{Z^{-1} - Z^{X}}\ , \qquad X \geq 0,\ K \in \kappa\ . \tag{2.6}$$

Then, by (2.1) and (2.3),

$$C_k(X,K) = Z^{K+k} C(X,K) , \quad k \geq 0, \ X \geq K, \quad K \in \kappa . \tag{2.7}$$

Also, by (2.0) and Assumption 2.2, $X^*(K)$ is equivalently determined by

$$C(X^*(K), K) = \min\{C(X,K) : \text{integer } X \in [K, U(K)]\} . \tag{2.8}$$

<u>Lemma 2.2</u>. For $X \geq 0$, $C(X,\cdot)$ is nondecreasing over κ.

<u>Proof</u>. Let $K,K' \in \kappa$ with $K > K'$. By (2.2), $\phi(K) \geq \phi(K')$. Hence,

$$C(X,K) = \frac{\phi(K) + g(X)}{Z^{-1} - Z^X} \geq \frac{\phi(K') + g(X)}{Z^{-1} - Z^X} = C(X,K') . \qquad \square$$

<u>Definition 2.1</u>. The function g is <u>convex</u> over the interval $[a,b]$ if $g(\mu X + (1-\mu)X') \leq \mu g(X) + (1-\mu)\, g(X')$ for all real $X,X' \in [a,b]$, $0 \leq \mu \leq 1$. The function g is <u>concave</u> over the interval $[a,b]$ if $-g$ is convex over the interval $[a,b]$. The function g is <u>discrete convex</u> over the interval $[a,b]$ if $g(\mu X + (1-\mu)X') \leq \mu g(X) + (1-\mu)\, g(X')$ for all integers $X,X' \in [a,b]$, $0 \leq \mu \leq 1$. The function g is <u>discrete concave</u> over the interval $[a,b]$ if $-g$ is discrete convex over $[a,b]$.

Recall that $g(0)$ represents the cost of a permanent expansion of size 1. Hence, a restriction of g to convexity over $[0,\infty)$ does <u>not</u> preclude the existence of a fixed (or "setup") charge in the expansion cost function. Notice that if g is convex (concave), then g is discrete convex (discrete concave).

2.2. Computing $X^*(K)$: Policy Improvement for General Discrete g

In this section, the determination of an optimal permanent expansion size $X^*(K) + 1$, for a fixed value $K \in \kappa$, is considered for arbitrary nonnegative, nondecreasing functions g. The approach used is based on Howard's well-known Policy Improvement Technique of dynamic programming [1]. In the context of the problem at hand, the Policy Improvement result can be paraphrased as follows for fixed $K \in \kappa$:

> "Using expansion size $X+1$ is more costly than using expansion size $X'+1$ if, and only if, using $X+1$ is more costly than using $X'+1$ for the first expansion and $X+1$ thereafter."

Theorem 2.5 is the precise statement of this result.

Definition 2.2. Let $G_k(K)$, $k \geq 0$, denote the expected discounted operating costs until the first permanent expansion, when the initial spares level is $k \geq 0$ (and no temporary facilities are initially in use) and the temporary facilities usage limit is $K \in \kappa$.

Lemma 2.3. For $K \in \kappa$,

$$C_k(X,K) = G_k(K) + (g(X) + C_{X-K}(X,K))\, Z^{K+k+1}, \qquad k \geq 0,\ X \geq K .$$

Proof. Starting from spares level k, the first expansion occurs whenever the net demand first increases by an amount $K+k+1$; this time is given by T_{K+k+1} and $Z^{K+k+1} = E[e^{-rT_{K+k+1}}]$. The expected discounted costs

over the time interval $[0, T_{K+k+1})$ are $G_k(K)$. At time T_{K+k+1}, an expansion is undertaken at cost $g(X)$ and the expected discounted costs thereafter are $C_{X-K}(X,K)$, relative to T_{K+k+1}. Hence,

$$\begin{aligned} C_k(X,K) &= G_k(K) + E[(g(X) + C_{X-K}(X,K))\, e^{-rT_{K+k+1}}] \\ &= G_k(K) + (g(X) + C_{X-K}(X,K))\, Z^{K+k+1} \quad . \qquad \square \end{aligned}$$

Lemma 2.4. For $K \in \kappa$,

$$C_k(X,K) = G_k(K) + \frac{Z^{K+k+1}}{1-Z^{X+1}} (g(X) + G_{X-K}(K)) \ , \qquad k \geq 0,\ X \geq K \ .$$

Proof. From the previous lemma,

$$C_{X-K}(X,K) = G_{X-K}(X,K) + (g(X) + C_{X-K}(X,K))\, Z^{X+1} \quad .$$

Iterating the above equation yields

$$\begin{aligned} C_{X-K}(X,K) = G_{X-K}(X,K) &+ (g(X) + G_{X-K}(X,K)) \sum_{i=1}^{\ell-1} Z^{i(X+1)} \\ &+ (g(X) + C_{X-K}(X,K))\, Z^{\ell(X+1)}, \qquad \ell = 1, 2, \ldots \ . \end{aligned}$$

Letting $\ell \to \infty$ in the above equation gives

$$C_{X-K}(X,K) = G_{X-K}(X,K) + (g(X) + G_{X-K}(X,K)) \sum_{i=1}^{\infty} Z^{i(X+1)} \quad .$$

Substituting this expression into that of the previous lemma yields

$$C_k(X,K) = G_k(K) + (g(X) + G_{X-K}(K))\,(1 + \sum_{i=1}^{\infty} Z^{i(X+1)})\, Z^{K+k+1}$$

$$= G_k(K) + Z^{K+k+1}(g(X) + G_{X-K}(K)) \sum_{i=0}^{\infty} (Z^{(X+1)})^i$$

$$= G_k(K) + \frac{Z^{K+k+1}}{1-Z^{X+1}}\,(g(X) + G_{X-K}(K)) \quad . \qquad \square$$

Theorem 2.5. (Policy Improvement). For $X,X' \geq K \in \kappa$,

$$C(X',K) \begin{Bmatrix} < \\ = \\ > \end{Bmatrix} C(X,K) \Longleftrightarrow g(X') + C_{X'-K}(X,K) \begin{Bmatrix} < \\ = \\ > \end{Bmatrix} g(X) + C_{X-K}(X,K) \quad .$$

Proof. We shall prove the theorem using the $<$ inequality; the other two cases can be proved in an analogous manner.

We shall first prove sufficiency ($\Leftarrow$). Let $h = Z^{X'+1}$. By Lemma 2.3 and the hypothesis,

$$C_{X'-K}(X,K) = G_{X'-K}(K) + (g(X) + C_{X-K}(X,K))h$$

$$> G_{X'-K}(K) + (g(X') + C_{X'-K}(X,K))h \quad . \qquad (2.9)$$

Iterating (2.9) $\ell+1$ times yields

$$C_{X'-K}(X,K) > G_{X'-K}(K) + (g(X') + G_{X'-K}(X,K)) \sum_{i=1}^{\ell} h^i$$

$$+ (g(X') + C_{X'-K}(X,K))\, h^{\ell+1}, \quad \ell = 0, 1, \ldots \quad . \qquad (2.10)$$

Since (2.9) is applied at each iteration, the righthand side of (2.10) decreases with ℓ. Hence, strict inequality is maintained in the limit, yielding as $\ell \to \infty$ that

$$C_{X'-K}(X,K) > G_{X'-K}(K) + \frac{h}{1-h}(g(X') + G_{X'-K}(K)) = C_{X'-K}(X',K) ,$$

where the equality follows from Lemma 2.4. By (2.7), the above inequality implies that $C(X,K) > C(X',K)$.

The proof of necessity ($\Rightarrow$) is by contradiction. Suppose that $C(X',K) < C(X,K)$, but contrary to the theorem,

$$g(X') + C_{X'-K}(X,K) \geq g(X) + C_{X-K}(X,K) \quad . \tag{2.11}$$

Let $h = Z^{X'+1}$. By Lemma 2.3 and the above inequality,

$$\begin{aligned} C_{X'-K}(X,K) &= G_{X'-K}(K) + (g(X) + C_{X-K}(X,K))h \\ &\leq G_{X'-K}(K) + (g(X') + C_{X'-K}(X,K))h \quad . \end{aligned}$$

Iterating this expression to the limit gives

$$C_{X'-K}(X,K) \leq G_{X'-K}(K) + (g(X') + G_{X'-K}(K))h/(1-h) = C_{X'-K}(X',K) ,$$

which by (2.7), implies that $C(X,K) \leq C(X',K)$. Since this contradicts the hypothesis, we must conclude that the supposition (2.11) is false, which completes the proof. □

Using (2.7), the Policy Improvement Theorem can be restated as follows.

Corollary 2.6. For $X, X' \geq K \in \kappa$,

$$C(X',K) \begin{Bmatrix} < \\ = \\ > \end{Bmatrix} C(X,K) \Longleftrightarrow g(X') - g(X) \begin{Bmatrix} < \\ = \\ > \end{Bmatrix} C(X,K)\ (Z^{X}-Z^{X'}) \ .$$

Proof.

$$C(X',K) \begin{Bmatrix} < \\ = \\ > \end{Bmatrix} C(X,K) \Longleftrightarrow g(X') - g(X) \begin{Bmatrix} < \\ = \\ > \end{Bmatrix} C_{X-K}(X,K) - C_{X'-K}(X,K)$$

$$= C(X,K)\ (Z^{K+X-K} - Z^{K+X'-K})$$

$$= C(X,K)\ (Z^{X} - Z^{X'}) \ . \qquad \square$$

Corollary 2.7. Let $X \geq K \in \kappa$. For $0 \leq x \leq U(K)-X$,

$$C(X+x,K) \begin{Bmatrix} < \\ = \\ > \end{Bmatrix} C(X,K) \Longleftrightarrow g(X+x) - g(X) \begin{Bmatrix} < \\ = \\ > \end{Bmatrix} C(X,K)\ Z^{X}(1-Z^{x}) \qquad (2.12)$$

For $0 \leq x \leq X-K$,

$$C(X-x,K) \begin{Bmatrix} < \\ = \\ > \end{Bmatrix} C(X,K) \Longleftrightarrow g(X) - g(X-x) \begin{Bmatrix} > \\ = \\ < \end{Bmatrix} C(X,K)\ Z^{X}(Z^{-x}-1) \ . \qquad (2.13)$$

Proof. (2.12) follows from Corollary 2.6 upon setting $X' = X+x$ and (2.13) follows likewise upon setting $X' = X-x$. □

Using Corollary 2.7, the Policy Improvement Algorithm for finding $X^*(K)$ can be stated in two different ways. Initializing the algorithm at the beginning of the interval $[K, U(K)]$ and using (2.12) gives:

Algorithm C_1 (Policy Improvement).

(1) $X_0 = K$; $i = 0$.

(2) Compute $C(X_i,K)$.

(3) Find $x' = \min\{x : g(X_i+x) - g(X_i) < C(X_i,K)\, Z^{X_i}(1-Z^x),\ 0 < x < U(K)-X_i\}$.
If x' does not exist, then $X^*(K) = X_i$ (stop).

(4) $X_{i+1} = X_i+x$; $i \leftarrow i+1$; go to step (2).

The sequence $\{X_i\}$ produced by Algorithm C_1 is increasing and $\{C(X_i,K)\}$ is decreasing. Since each (integer) $X_i \in [K, U(K)]$, the algorithm is finite.

Initializing the algorithm at the end of the interval $[K, U(K)]$ and using (2.13) gives:

Algorithm C_2 (Policy Improvement).

(1) $X_0 = U(K)$; $i = 0$.

(2) Compute $C(X_i,K)$.

(3) Find $x' = \min\{x : g(X_i) - g(X_i-x) > C(X_i,K)\, Z^{X_i}(Z^{-x}-1),\ 0 < x \leq X_i-K\}$.
If x' does not exist, then $X^*(K) = X_i$ (stop).

(4) $X_{i+1} = X_i - x'$; $i \leftarrow i+1$; go to step (2).

The sequence $\{X_i\}$ produced by Algorithm C_2 is decreasing, as is $\{C(X_i,K)\}$. Since each (integer) $X_i \in [K, U(K)]$, the algorithm is finite.

It should be noted that Corollary 2.6 can also be used to state an algorithm that permits an arbitrary initial choice $X_0 \in [K, U(K)]$. In this case, the sequence $\{C(X_i,K)\}$ will again be decreasing, and hence the algorithm is finite since $\{X_i\}$ will be a nonrepeating sequence on a finite set. However, the sequence $\{X_i\}$ will, in general, not be monotone; this is often impractical. For this reason, Algorithms C_1 and C_2 would be preferred in practice.

While Policy Improvement provides finite algorithms, it is, in the worst case, no better than an exhaustive enumeration of the interval $[K, U(K)]$. However, the Policy Improvement Theorem can be used to obtain conditions on g which allow more efficient search techniques. The next section addresses this subject.

2.3. Computing $X^*(K)$: Unimodal Properties, Interval Bisection and Monotonicity for Discrete g.

In this section, the Policy Improvement Theorem is used to derive conditions on g which abet more efficient search techniques. General conditions are derived for g which will guarantee that $C(\cdot,K)$ is unimodal for fixed $K \in \kappa$.

Definition 2.3. For fixed $K \in \kappa$, $C(\cdot,K)$ is unimodal over the interval $[K, U(K)]$ if there exists $X' \in [K, U(K)]$ such that:

(i) $C(X,K) \geq C(X+1, K)$, $K \leq X < X'$; and

(ii) $C(X,K) \leq C(X+1, K)$, $X' \leq X < U(K)$.

Definition 2.4. For $K \in \kappa$, let $\chi_1(K)$ and $\chi_2(K)$ denote the sets given by:

$$\chi_1(K) = \{X : g(X+1) - g(X) > C(X,K)\ Z^X(1-Z),\ K \leq X < U(K)\} ,$$

$$\chi_2(K) = \{X : g(X) - g(X-1) < C(X,K)\ Z^X(Z^{-1}-1),\ K < X \leq U(K)\} .$$

Let $X_1(K)$ and $X_2(K)$ be given by

$$X_1(K) = \begin{cases} U(K), & \text{if } \chi_1(K) = \emptyset \\ \min(X : X \in \chi_1(K)), & \text{if } \chi_1(K) \neq \emptyset \end{cases}$$

$$X_2(K) = \begin{cases} K, & \text{if } \chi_2(K) = \emptyset \\ \max(X : X \in \chi_2(K)), & \text{if } \chi_2(K) \neq \emptyset \end{cases}$$

The following lemma provides alternative definitions for the above two sets.

Lemma 2.8. For $K \in \kappa$ and $X = [K, U(K)]$,

$$X \in \chi_1(K) \iff C(X,K) < C(X+1, K) ,$$

and

$$X \in \chi_2(K) \iff C(X-1, K) > C(X,K) .$$

Proof. Follows from Definition 2.4 and Corollary 2.7. □

Lemma 2.9. For $K \in \kappa$,

$$K \leq X < X_1(K) \Rightarrow C(X,K) \geq C(X+1, K) \Rightarrow C(X,K) \geq C(X_1(K), K) ,$$

and

$$X_2(K) < X \leq U(K) \Rightarrow C(X-1,K) \leq C(X,K) \Rightarrow C(X_2(K), K) \leq C(X,K) .$$

Proof. From the previous lemma and Definition 2.4,

$$K \leq X < X_1(K) \Rightarrow X \notin \chi_1(K) \iff C(X,K) \geq C(X+1,K)$$

$$X_2(K) < X \leq U(K) \Rightarrow X \notin \chi_2(K) \iff C(X-1, K) \leq C(X,K) .$$ □

Thus, if $\chi_1(K) \neq \emptyset$, then $X_1(K)$ is the first point of increase for $C(\cdot,K)$ over $[K, U(K)]$ (i.e., $C(X_1(K)+1, K) > C(X_1(K), K)$). Similarly, if $\chi_2(K) \neq \emptyset$, then $X_2(K)$ is the last point of decrease for $C(\cdot,K)$ over $[K, U(K)]$ (i.e., $C(X_2(K)-1, K) > C(X_2(K), K)$).

Theorem 2.10 (Monotonicity). $X_1(\cdot)$ and $X_2(\cdot)$ are each nondecreasing over κ.

Proof. Let $K, K' \in \kappa$ with $K < K'$. If $X_1(K') \geq U(K)$, then the result for $X_1(\cdot)$ follows trivially. Therefore, assume that $X_1(K') \leq U(K)-1$ $\Rightarrow X_1(K') \leq U(K')-1 \Rightarrow \mathcal{X}_1(K') \neq \emptyset$. Hence, using Definition 2.4 and Lemma 2.2,

$$\begin{aligned} g(X_1(K')+1) - g(X_1(K')) &> C(X_1(K'), K')\, Z^{X_1(K')}(1-Z) \\ &\geq C(X_1(K'), K)\, Z^{X_1(K')}(1-Z) , \end{aligned}$$

which implies that $X_1(K') \in \mathcal{X}_1(K)$, since $X_1(K') \in [K', U(K)-1]$ $[K, U(K)-1]$. Therefore, $\mathcal{X}_1(K) \neq \emptyset$ and $X_1(K') \geq \min\{X : X \in \mathcal{X}_1(K)\} = X_1(K)$.

If $X_2(K) \leq K'$, then the result for $X_2(\cdot)$ follows trivially. Therefore, assume that $X_2(K) \geq K'+1 \Rightarrow X_2(K) \geq K+1 \Rightarrow \mathcal{X}_2(K) \neq \emptyset$. Hence, using Definition 2.4 and Lemma 2.2,

$$\begin{aligned} g(X_2(K)) - g(X_2(K) - 1) &< C(X_2(K), K)\, Z^{X_2(K)}(Z^{-1}-1) \\ &\leq C(X_2(K), K')\, Z^{X_2(K)}(Z^{-1}-1) , \end{aligned}$$

which implies that $X_2(K) \in \mathcal{X}_2(K')$, since $X_2(K) \in [K'+1, U(K)]$ $\subseteq [K'+1, U(K')]$. Therefore, $\mathcal{X}_2(K') \neq \emptyset$ and $X_2(K) \leq \max\{X : X \in \mathcal{X}_2(K')\}$ $= X_2(K')$. □

The above theorem is independent of the functional form for g and, in that context, it is worthy of restatement. The Monotonicity Theorem states that within κ: (a) points of first increase for $C(\cdot,K)$ over $[K, U(K)]$ are nondecreasing in K; and (b) points of last decrease for $C(\cdot,K)$ over $[K, U(K)]$ are also nondecreasing in K.

The next series of propositions derive conditions on g that lead to general unimodal results.

<u>Definition 2.5</u>. Let $K \in \kappa$. g will be said to satisfy <u>property</u> Λ_K if

$$g(X+1) - g(X) \geq Z(g(X) - g(X-1)) , \qquad K+1 \leq X \leq U(K)-1$$

g will be said to satisfy <u>property</u> Λ if g satisfies property Λ_K for all $K \in \kappa$.

The above definition is dependent on the value Z, as determined by the probability parameters (λ,p,q) in (2.3). However, the following result is independent of these parameters.

<u>Lemma 2.11</u>. Let $K \in \kappa$. If g is discrete convex over $[K, U(K)]$, then g satisfies property Λ_K. If g is discrete convex over $[0,\infty)$, then g satisfies property Λ.

Proof. g discrete convex over $[K, U(K)]$ implies that

$$g(X+1) - g(X) \geq g(X) - g(X-1) \geq Z(g(X) - g(X-1)), \quad K+1 \leq X \leq U(K)-1 ,$$

since $0 < Z < 1$. If g is discrete convex over $[0,\infty)$, then g is discrete convex over $[K, U(K)]$, for all $K \in \kappa$. □

The next two lemmas list equivalent definitions for the property Λ_K.

Lemma 2.12: Let $K \in \kappa$. g satisfies property Λ_K if and only if

$$g(X+x) - g(X) \geq \frac{1-Z^x}{1-Z} (g(X+1) - g(X)) ,$$

$$1 \leq x \leq U(K)-X, \; K \leq X \leq U(K)-1 . \tag{2.14}$$

Proof. To prove necessity ($\Rightarrow$), suppose g satisfies property Λ_K. Then,

$$\begin{aligned} g(X+i) - g(X+i-1) &\geq Z(g(X+i-1) - g(X+i-2)) \\ &\geq \cdots \\ &\geq Z^{i-1}(g(X+1) - g(X)) . \end{aligned}$$

Hence,

$$g(X+x) - g(X) = \sum_{i=1}^{x} (g(X+i) - g(X+i-1))$$

$$\geq (g(X+1) - g(X)) \sum_{i=1}^{x} Z^{i-1}$$

$$= (g(X+1) - g(X)) \frac{1-Z^x}{1-Z} \quad .$$

To prove sufficiency ($\Leftarrow$), let $x = 2$ in (2.14). Then,

$$g(X+2) - g(X) \geq \frac{1-Z^2}{1-Z} (g(X+1) - g(X))$$

$$= (1+Z) (g(X+1) - g(X))$$

$$\Rightarrow g(X+2) - g(X+1) \geq Z(g(X+1) - g(X)) \quad .$$ □

<u>Lemma 2.13</u>. Let $K \in \kappa$. g satisfies property Λ_K if and only if

$$g(X) - g(X-x) \leq \frac{Z^{-x}-1}{Z^{-1}-1} (g(X) - g(X-1)) \ ,$$

$$1 \leq x \leq X-K, \ K+1 \leq X \leq U(K) \quad . \qquad (2.15)$$

<u>Proof</u>. To prove necessity ($\Rightarrow$), suppose g satisfies property Λ_K. Then,

$$g(X-i) - g(X-i-1) \leq Z^{-1}(g(X-i+1) - g(X-i))$$

$$\leq \cdots$$

$$\leq Z^{-i}(g(X) - g(X-1)) \quad .$$

Hence,

$$g(X) - g(X-x) = \sum_{i=0}^{x-1} (g(X-i) - g(X-i-1))$$

$$\leq (g(X) - g(X-1)) \sum_{i=0}^{x-1} Z^{-i}$$

$$= (g(X) - g(X-1)) \frac{Z^{-x}-1}{Z^{-1}-1} \quad .$$

To prove sufficiency ($\Leftarrow$), let $x = 2$ in (2.15). Then,

$$g(X) - g(X-2) \leq \frac{Z^{-2}-1}{Z^{-1}-1} (g(X) - g(X-1))$$

$$= (Z^{-1}+1) \ (g(X) - g(X-1))$$

$$\Rightarrow g(X-1) - g(X-2) \leq Z^{-1}(g(X) - g(X-1))$$

$$\Rightarrow g(X) - g(X-1) \geq Z(g(X-1) - g(X-2)) \quad . \qquad \square$$

Lemma 2.14. Let $K \in \kappa$. If $X_1(K) \geq X_2(K)$, then

(i) $C(\cdot,K)$ is unimodal over $[K, U(K)]$, and

(ii) $C(X,K) = C(X^*(K), K)$, $X \in [X_2(K), X_1(K)]$.

Proof. By Lemma 2.9, $X_1(K) \geq X_2(K)$ implies

$$C(X,K) \geq C(X+1, K) , \qquad K \leq X < X_2(K)$$

$$C(X,K) = C(X+1, K) , \qquad X_2(K) \leq X < X_1(K)$$

$$C(X,K) \leq C(X+1, K) , \qquad X_1(K) \leq X < U(K) \quad . \qquad \square$$

Theorem 2.15 (Unimodality). Let $K \in \kappa$. If g satisfies property Λ_K, then

(i) $C(\cdot,K)$ is unimodal over $[K, U(K)]$,

(ii) $C(X,K) = C(X^*(K), K)$, $X \in [X_2(K), X_1(K)]$, and

(iii) $C(X,K) > C(X^*(K), K)$, $X \in [K, U(K)] \sim [X_2(K), X_1(K)]$.

Proof. By Definition 2.4,

$$g(X_1(K) + 1) - g(X_1(K)) > C(X_1(K), K)\, Z^{X_1(K)}(1-Z)\ , \quad X_1(K) \neq \emptyset\ .$$

Combining this with inequality (2.14) gives

$$g(X_1(K) + x) - g(X_1(K)) > C(X_1(K), K)\, Z^{X_1(K)}(1-Z^x)\ ,$$

$$1 \leq x \leq U(K)-X_1(K),\ X_1(K) \neq \emptyset\ .$$

Hence, by Corollary 2.7,

$$C(X,K) > C(X_1(K),K),\ X_1(K) < X \leq U(K),\ X_1(K) \neq \emptyset\ . \tag{2.16}$$

Similarly, by Definition 2.4,

$$g(X_2(K)) - g(X_2(K)-1) < C(X_2(K),K)\, Z^{X_2(K)}(Z^{-1}-1)\ , \quad X_2(K) \neq \emptyset\ .$$

Combining this with inequality (2.15) gives

$$g(X_2(K)) - g(X_2(K) - x) < C(X_2(K), K)\, Z^{X_2(K)}(Z^{-x}-1)\ ,$$

$$1 \le x \le X_2(K)-K,\ \chi_2(K) \ne \emptyset\ .$$

Hence, by Corollary 2.7,

$$C(X_2(K),K) < C(X,K), \quad K \le X < X_2(K),\ \chi_2(K) \ne \emptyset\ . \tag{2.17}$$

We now consider four possible cases.

Case 1: $\chi_1(K) = \emptyset$ and $\chi_2(K) = \emptyset$. Then $X_2(K) = K \le U(K) = X_1(K)$ and (i)-(ii) follow from the previous lemma. (iii) is null for this case.

Case 2: $\chi_1(K) \ne \emptyset$ and $\chi_2(K) = \emptyset$. Then $X_2(K) = K \le X_1(K)$ and (i)-(ii) follow from the previous lemma. (iii) follows from (2.16).

Case 3: $\chi_1(K) = \emptyset$ and $\chi_2(K) \ne \emptyset$. Then $X_2(K) \le U(K) = X_1(K)$ and (i)-(ii) follow from the previous lemma. (iii) follows from (2.17).

Case 4: $\chi_1(K) \ne \emptyset$ and $\chi_2(K) \ne \emptyset$. Suppose $X_1(K) < X_2(K)$. Then (2.16) and (2.17) together imply that $C(X_2(K), K) > C(X_1(K), K) > C(X_2(K), K)$, a contradiction. Hence, $X_1(K) \ge X_2(K)$ and (i)-(ii) follow from the previous lemma. (iii) follows from (2.16) and (2.17). □

By (ii) and (iii) of the above theorem, $X_1(K)$ and $X_2(K)$ are respectively the greatest and least elements in the minimizing set for $C(\cdot,K)$ over $[K, U(K)]$. Hence, whenever g satisfies property Λ_K, $X_2(K)+1$ is the <u>least optimal expansion size</u> and $X_1(K)+1$ is the <u>greatest optimal expansion size</u>. Combining this observation with the Monotonicity Theorem gives

<u>Corollary 2.16</u> (Monotonicity). If g satisfies property Λ, then (i)-(iii) of Theorem 2.15 hold for all $K \in \kappa$, and the least and greatest optimal expansion sizes ($1+X_2(\cdot)$ and $1+X_1(\cdot)$, respectively) are each nondecreasing over κ.

<u>Proof</u>. The result follows directly from Theorems 2.10 and 2.15. □

<u>Corollary 2.17</u>. Let $K \in \kappa$; if g is discrete convex over $[K, U(K)]$, then (i)-(iii) of Theorem 2.15 follow. If g is discrete convex over $[0,\infty)$, then (i)-(iii) of Theorem 2.15 hold for all $K \in \kappa$ and the least and greatest optimal expansion sizes ($1+X_2(\cdot)$ and $1+X_1(\cdot)$, respectively) are each nondecreasing over κ.

<u>Proof</u>. The result follows directly from Theorem 2.15, Corollary 2.16, and Lemma 2.11. □

The above propositions detail conditions on g which guarantee that $C(\cdot,K)$ is unimodal over $[K, U(K)]$. In these cases, $X^*(K)$ can be found by using Interval Bisection ("Binary Search"):

Algorithm C_3 (Interval Bisection, given Λ_K).

(1) $X_\ell = K$; $X_u = U(K)$.

(2) If $X_\ell = X_u$, then $X^*(K) = X_\ell = X_u$ (stop).

(3) $X = [(X_\ell + X_u)/2]$.

(4) If $g(X+1) - g(X) \geq C(X,K)\ Z^X(1-Z)$, then reset $X_\ell = X$; otherwise, reset $X_u = X+1$.

(5) Go to step (2).

Using Algorithm C_3, the search interval is essentially halved at each iteration and no more than $1 + \log_2(U(K) - K)$ iterations are required. Note that by Corollary 2.7, the "if" condition of step (4) is equivalent to: "If $C(X,K) \leq C(X+1, K)$". Thus, the iterate X_ℓ is only reset to X if $C(X',K) > C(X'+1,K)$ for $X' \leq X-1$. Hence, the algorithm computes $X_2(K)$, the least optimal expansion size (less unity). By removing the equality portion of the "if" condition in step (4), the algorithm will alternatively compute $X_1(K)$, the greatest optimal expansion size (less unity). Obviously, the algorithm can be easily modified to compute both $X_1(K)$ and $X_2(K)$, thereby giving the entire set of optimal expansion sizes for $K : [X_2(K)+1, X_1(K)+1]$.

In view of Corollary 2.16, some additional efficiency can be realized in Algorithm C_3 whenever it is used to recursively compute $X^*(\cdot)$ over κ. To illustrate, suppose that K' and K are consecutive elements

of κ with $K' < K$. Then $X^*(K')$ is a lower bound on $X^*(K)$ and additional efficiency may be realized by alternatively initializing the lower iterate $X_\ell = \max\{K, X^*(K')\}$ in order to compute $X^*(K)$.

The final theorem of this section provides a general result that may be used to construct more efficient search algorithms for a number of other cases where Property Λ_K cannot be ascertained. It demonstrates that further restriction of the search interval may be possible, given known bounds on the marginal costs (e.g., g a step function). This result can be used for arbitrary g is conjunction with Algorithm C_1 of the previous section.

<u>Theorem 2.18</u>. Let $K \in \kappa$. Suppose $0 < \xi_1 \leq g(X+1) - g(X) \leq \xi_2$, for $X \in [K, U(K)]$. *Then* $X^*(K) \in [X_1(K), U_1(K)]$, where $U_1(K) = \min\{U(K), X_1(K) + [\xi_2 \xi_1^{-1}(1-Z)^{-1}]\}$.

<u>Proof</u>. By Lemma 2.9, it sufficies to show that $X > U_1(K)$ implies $X \neq X^*(K)$ whenever $U_1(K) < U(K)$. When $U_1(K) < U(K)$,

$$X > U_1(K) \Rightarrow X - X_1(K) = x \geq [\xi_2 \xi_1^{-1}(1-Z)^{-1}] + 1$$

$$\begin{aligned} \Rightarrow g(X) - g(X_1(K)) &= g(X_1(K) + x) - g(X_1(K)) \\ &\geq \xi_1 x \\ &\geq \xi_1(\xi_2 \xi_1^{-1}(1-Z)^{-1}) \\ &\geq \xi_2 \frac{1-Z^x}{1-Z} \\ &\geq \frac{1-Z^x}{1-Z}\,(g(X_1(K)+1) - g(X_1(K)))\ . \end{aligned}$$

By Corollary 2.7, this implies that $C(X,K) > C(X_1(K), K)$. □

2.4. Computing $X^*(K)$: Function Iteration and Monotonicity for Concave g

In this section, the determination of $X^*(K)$ is considered for the important case where g is concave. Assumption 2.2, introduced in Section 2.1, will be retained. However, Assumption 2.1 will be strengthened:

Assumption 2.1': g is positive, increasing and continuously differentiable.

Under Assumption 2.1', $C(\cdot,K)$ is continuous for each fixed value $K \geq 0$. Unlike previous sections of this chapter, X will be treated as a continuous variable here. However, $X^*(K) + 1$ will continue to denote the optimal integer expansion size. Notice that the positivity of g is not impractical since $g(0)$ has been defined as the cost for a permanent expansion of size 1.

Definition 2.5.

$$f(X,K) = \frac{1}{\ln Z^{-1}} \ln\left(1 + \frac{g(X) + \phi(K)}{g'(X)} \ln Z^{-1}\right) - 1, \qquad X \geq 0,\ K \in \kappa.$$

The following lemma demonstrates that the function f is well-defined under Assumption 2.1'.

Lemma 2.19. $f(X,K) > -1$, $X \geq 0$, $K \in \kappa$.

Proof. Let $K \in \kappa$ and $X \geq 0$. Since $0 < Z < 1$, $\ln Z^{-1} > 0$. By Assumption 2.1' and (2.2), $(g(X) + \phi(K))/g'(X) > 0$. Hence,

$$\frac{1}{\ln Z^{-1}} \ln\left(1 + \frac{g(X) + \phi(K)}{g'(X)} \ln Z^{-1}\right) > 0 \;. \qquad \square$$

Under Assumption 2.1', $f(\cdot,K)$ is continuous for each fixed $K \in \kappa$. The significant of f is given by:

Theorem 2.20.

$$\operatorname{sgn}\{\frac{\partial}{\partial X} C(X,K)\} = \operatorname{sgn}\{X - f(X,K)\} \;, \qquad X \geq 0, \quad K \in \kappa \quad .$$

Proof. Using (2.6),

$$\frac{\partial}{\partial X} C(X,K) = \frac{g'(X)\, Z^{-1} - Z^X (g'(X) + \ln Z^{-1}(g(X) + \phi(K)))}{(Z^{-1} - Z^X)^2} \;.$$

Rearranging terms in the above numerator gives

$$\frac{\partial}{\partial X} C(X,K) \begin{Bmatrix} > \\ = \\ < \end{Bmatrix} 0 \iff Z^X \begin{Bmatrix} < \\ = \\ > \end{Bmatrix} \frac{g'(X)\, Z^{-1}}{g'(X) + \ln Z^{-1}(g(X) + \phi(K))} \;.$$

Note that the above right hand side is positive by Assumption 2.1' and (2.2). Taking logarithms yields

$$\frac{\partial}{\partial X} C(X,K) \begin{Bmatrix} > \\ = \\ < \end{Bmatrix} 0 \iff X \begin{Bmatrix} > \\ = \\ < \end{Bmatrix} f(X,K) \;. \qquad \square$$

Thus, if $X < +\infty$ is a global minima of $C(\cdot,K)$ over $[0,\infty)$, then X must be a fixed point of $f(\cdot,K)$; that is, $X = f(X,K)$. The behavior of f with respect to K is given by the following lemma.

Lemma 2.21. $f(X,\cdot)$ is nondecreasing over κ, $X \geq 0$.

Proof. Let $K, K' \in \kappa$ with $K' < K$. Let $X \geq 0$. By (2.2) and positivity,

$$1 + \frac{g(X) + \phi(K')}{g'(X)} \ln Z^{-1} \leq 1 + \frac{g(X) + \phi(K)}{g'(X)} \ln Z^{-1} .$$

Taking logarithms and multiplying both sides by $(\ln Z^{-1})^{-1} > 0$ yields $f(X,K') + 1 \leq f(X,K) + 1$. □

When g is concave, f is also monotone with regard to X.

Lemma 2.22. If g is concave, then $f(\cdot,K)$ is increasing, $K \in \kappa$.

Proof. Let $K \in \kappa$, $X \geq 0$ and $\epsilon > 0$. When g is concave, g' is nonincreasing. Hence

$$g(X) \leq g(X+\epsilon) - g'(X+\epsilon)\epsilon \quad \text{and} \quad g'(X) \geq g'(X+\epsilon) > 0$$

$$\Rightarrow \frac{g(X)}{g'(X)} \leq \frac{g(X+\epsilon) - g'(X+\epsilon)\epsilon}{g'(X)} < \frac{g(X+\epsilon)}{g'(X)} \leq \frac{g(X+\epsilon)}{g'(X+\epsilon)}$$

$$\Rightarrow \frac{g(X) + \phi(K)}{g'(X)} < \frac{g(X+\epsilon) + \phi(K)}{g'(X+\epsilon)}$$

$$\Rightarrow v \equiv \left(\frac{1 + \frac{g(X+\epsilon) + \phi(K)}{g'(X+\epsilon)} \ln Z^{-1}}{1 + \frac{g(X) + \phi(K)}{g'(X)} \ln Z^{-1}} \right) > 1$$

$$\Rightarrow \ln v > 0$$

$$\Rightarrow f(X+\epsilon,\ K) - f(X,K) = \frac{1}{\ln Z^{-1}} \ln v > 0 \ . \qquad \square$$

<u>Definition 2.6</u>. For $K \in \kappa$, <u>condition</u> Σ_k will be said to hold if g is concave and if there exists $\epsilon_k > 0$ such that

(Σ_K^1): $\quad f(X,K) > X$, $\quad X \in (0,\ \epsilon_K]$

and

(Σ_K^2): $\quad f(\cdot,K)$ is concave over $[\epsilon_K,\ \infty)$.

<u>Condition</u> Σ will be said to hold if Σ_K holds for all $K \in \kappa$.

Later in this section, it will be demonstrated that condition Σ holds for several major classes of concave functions. The following lemma lists several properties of the above conditions.

Lemma 2.23. Suppose g is concave.

(i) If $f(\cdot,K)$ is concave, then (Σ_K^2) is satisfied, $K \in \kappa$.

(ii) If $(\Sigma_{K'}^1)$ is satisfied, for some $K' \in \kappa$, then (Σ_K^1) is satisfied for all $K \geq K'$, $K \in \kappa$.

(iii) If $f(0,0) > 0$, then (Σ_K^1) is satisfied for all $K \in \kappa$.

(iv) If (Σ_K^2) is satisfied for all $K \in \kappa$ and $f(0,0) > 0$, then Σ holds.

(v) If $f(\cdot,K)$ is concave for all $K \in \kappa$ and $f(0,0) > 0$, then Σ holds.

Proof. $f(\cdot,K)$ concave over $[0,\infty)$ obviously satisfies (Σ_K^2), so (i) follows trivially. By Lemma 2.21, $f(X,K) \geq f(X,K')$ for $K \geq K'$; $K', K \in \kappa$. Hence, $f(X,K') > X$, $X \in (0, \epsilon_{K'}] \Rightarrow f(X,K) > X$, $X \in (0, \epsilon_{K'}]$, $K \geq K'$, $K \in \kappa$; this proves (ii). By continuity, $f(0,0) > 0$ implies $\exists\, \epsilon_0 > 0$ such that $f(X,0) > X$ for $X \in [0, \epsilon_0]$; hence (iii) follows from (ii). (iv) follows from (iii). (v) follows from (i) and (iv). □

The next lemma implies that, in typical applications, f will be positive whenever g is concave (and hence, (Σ_K^1) will be satisfied for all $K \in \kappa$ -- see (iii) above).

Lemma 2.24. If $\frac{g'(0)}{g(0)} \leq 1 - \frac{\ln Z^{-1}}{2}$, then $f(0,0) > 0$.

Proof. By (2.2), $\phi(0) \geq 0$. Hence

$$\frac{g'(0)}{g(0)} \leq 1 - \frac{\ln Z^{-1}}{2} \Rightarrow \frac{2g'(0)}{\phi(0) + g(0)} + \ln Z^{-1} \leq 2$$

$$\Rightarrow 2 \; \frac{\phi(0) + g(0)}{2g'(0) + \ln Z^{-1}(\phi(0) + g(0))} \geq 1$$

$$\Rightarrow 2 \; \frac{\ln Z^{-1}(\phi(0) + g(0))/g'(0)}{2 + \ln Z^{-1}(\phi(0) + g(0))/g'(0)} \geq \ln Z^{-1} \; .$$

The natural logarithm expansion

$$\ln(1+v) = 2 \sum_{i=0}^{\infty} \left(\frac{v}{2+v}\right)^{2i+1}$$

for $v > 0$ thus gives

$$\ln\left(1 + \frac{\phi(0) + g(0)}{g'(0)} \ln Z^{-1}\right) > 2 \; \frac{\ln Z^{-1}(\phi(0) + g(0))/g'(0)}{2 + \ln Z^{-1}(\phi(0) + g(0))/g'(0)} \geq \ln Z^{-1}$$

$$\Rightarrow \frac{1}{\ln Z^{-1}} \ln\left(1 + \frac{\phi(0) + g(0)}{g'(0)} \ln Z^{-1}\right) > 1$$

$$\Rightarrow f(0,0) > 0 \; . \qquad \square$$

Lemma 2.25. If g is concave and $\frac{g'(0)}{g(0)} \leq 1 - \frac{\ln Z^{-1}}{2}$, then (Σ_K^1) is satisfied, $K \in \kappa$.

Proof. This result follows directly from Lemma 2.24 and Lemma 2.23,(iii).

$\square$

In practice, when g is concave, it is reasonable to expect the hypothesis of Lemma 2.24 to hold. This will now be demonstrated. Using (2.3), Z can be parameterized in (r, λ, p):

$$Z(r,\lambda,p) = \frac{(\lambda+r) - \sqrt{(\lambda+r)^2 - 4p(1-p)\lambda^2}}{2\lambda(1-p)} . \tag{2.18}$$

Although the proof of the following lemma is analytic, the result can also be obtained by probabilistic arguments using (2.5).

<u>Lemma 2.26</u>. Z decreases with r, increases with λ, and increases with p.

<u>Proof</u>. Denote

$$v = (\lambda+r) - \sqrt{(\lambda+r)^2 - 4p(1-p)\lambda^2} > 0 \quad .$$

Then

$$Z = \frac{(\lambda+r)-v^{1/2}}{2\lambda(1-p)} \Rightarrow v^{1/2} = (\lambda+r) - 2\lambda(1-p)Z$$

$$\Rightarrow v = (\lambda+r)^2 - 4\lambda(\lambda+r)(1-p)Z + 4\lambda^2(1-p)^2 Z^2 \quad .$$

These relations will be used extensively.

Taking the derivative of Z with respect to r and multiplying by $1 = v^{1/2} v^{-1/2}$ yields

$$\mathrm{sgn}(\frac{dZ}{dr}) = \mathrm{sgn}(v^{1/2} - (\lambda+r)) = \mathrm{sgn}(-2\lambda(1-p)Z) ,$$

so $dZ/dr < 0$.

Taking the derivative of Z with respect to λ and multiplying by $1 = v^{1/2}\, v^{-1/2}$ yields

$$
\begin{aligned}
\operatorname{sgn}\left(\frac{dZ}{d\lambda}\right) &= \operatorname{sgn}((\lambda+r)^2 - \lambda(\lambda+r) - rv^{1/2}) \\
&= \operatorname{sgn}((\lambda+r)^2 - \lambda(\lambda+r) - r((\lambda+r) - 2\lambda(1-p)Z)) \\
&= \operatorname{sgn}((\lambda+r)^2 - (\lambda+r)^2 + 2\lambda r(1-p)Z) \\
&= \operatorname{sgn}(2\lambda r(1-p)Z) ,
\end{aligned}
$$

so $dZ/d\lambda > 0$.

Recall (see (2.5)) that Z satisfies $\alpha Z^2 - Z + \beta = 0$. Taking the derivative of Z with respect to p and multiplying by $1 = v^{1/2}\, v^{-1/2}$ gives

$$
\begin{aligned}
\operatorname{sgn}\left(\frac{dZ}{dp}\right) &= \operatorname{sgn}(-v + (\lambda+r)\, v^{1/2} + 2\lambda^2(1-p)(1-2p)) \\
&= \operatorname{sgn}(-2\lambda(1-p)Z^2 + (\lambda+r)Z + \lambda(1-2p)) \\
&= \operatorname{sgn}\left(- \frac{2\lambda(1-p)}{\lambda+r} Z^2 + Z + \frac{\lambda(1-2p)}{\lambda+r}\right) \\
&= \operatorname{sgn}\left(\left(- \frac{\lambda(1-p)}{\lambda+r} Z^2 + Z - \frac{\lambda p}{\lambda+r}\right) + \frac{\lambda(1-p)}{\lambda+r} (1-Z^2)\right) \\
&= \operatorname{sgn}((-\alpha Z^2 + Z - \beta) + \beta(1-Z^2)) \\
&= \operatorname{sgn}(\beta(1-Z^2)) ,
\end{aligned}
$$

so $dZ/dp > 0$. $\square$

Lemma 2.27. $1 - \frac{\ln Z^{-1}}{2}$ decreases with r, increases with λ and increases with p.

Proof. $-\ln Z^{-1} = \ln Z$, so the result follows from Lemma 2.26. □

Suppose that

$$\frac{g'(0)}{g(0)} \leq 1 - \frac{\ln Z^{-1}(r', \lambda', p')}{2}$$

for a particular set of parameters (r', λ', p'). Then, by the preceeding lemma,

$$\frac{g'(0)}{g(0)} \leq 1 - \frac{\ln Z^{-1}(r,\lambda,p)}{2}$$

for $r \leq r'$, $\lambda \geq \lambda'$ and $p \geq p'$. Thus, given the ratio $g'(0)/g(0)$, it is possible to determine regions of the parameters (r, λ, p) over which the hypothesis of Lemma 2.24 is satisfied. Table 2.1 is a compilation of such regions for various upper bounds on the ratio $g'(0)/g(0)$. For example, the first line of the table states that if $g'(0)/g(0) \leq .1$, then $f(0,0) > 0$ for $r \leq .3$, $\lambda \geq 1$ and $p \geq .2$ (which, in view of the comments in [5], probably hold when the model is applicable).

Recall that $g(0)$ is the total cost for the first unit of capacity added. Thus, when the expansion cost is concave, $g'(0)$ is an upper bound on the marginal cost $(g(1) - g(0))$ for the second unit of capacity and

$\frac{g'(0)}{g(0)} \leq$	$r \leq$	$\lambda \geq$	$p \geq$
.1	.3	1	.2
.2	.3	1	.24
.3	.2	1	.26
.4	.2	1	.31
.5	.2	1	.36
.6	.2	10	.33
.7	.2	10	.38
.8	.2	10	.43
.9	.2	100	.5

TABLE 2.1: Regions where $f(0,0) > 0$

a lower bound on the marginal cost (over fixed cost) for the first unit of capacity. Hence, when g is concave,

$$\frac{g'(0)}{g(0)} = \frac{g'(0)}{\text{fixed costs + marginal cost of 1st unit added}}$$

$$\leq \frac{\text{marginal cost of 1st unit added}}{\text{fixed cost + marginal cost of 1st unit added}} .$$

Thus, when expansion costs are concave, the ratio $g'(0)/g(0)$ will certainly be less than unity. In practice, the fixed (or "setup") cost

is typically much greater than the marginal costs, so it is not unreasonable to expect $g'(0)/g(0)$ to be relatively small when expansion costs are concave ($g'(0)/g(0) \leq .1$ is not atypical). In summary, it seems reasonable to conclude that f will be positive (implying that (Σ_K^1) is satisfied for $K \in \kappa$) in most applications involving concave expansion costs.

Definition 2.7. For $K \in \kappa$, let $\chi(K)$ denote the fixed points of $f(\cdot,K)$ over $(0,\infty)$:

$$\chi(K) = \{X : X = f(X,K),\ X \in (0,\infty)\} \ .$$

If $\chi(K) \neq \emptyset$, let $x(K) = \min\{X : X \in \chi(K)\}$. If $\chi(K) = \emptyset$, let $x(K) = +\infty$.

Theorem 2.28. If Σ_K holds, then $|\chi(K)| \leq 1$, $K \in \kappa$.

Proof. Let $K \in \kappa$. By Definition 2.6, $\chi(K) \cap (0, \epsilon_K] = \emptyset$. Also, $f(\epsilon_K,K) \geq \epsilon_K$ and $f(\cdot,K)$ is concave over $[\epsilon_K,\infty)$. If $\chi(K) = \emptyset$, there is nothing to prove. Hence, suppose $\chi(K) \neq \emptyset$, which implies $x(K) < +\infty$. Let $X > x(K)$ be arbitrarily chosen. Let $\mu = (X-x(K))/(x-\epsilon_K)$. Then $0 < \mu < 1$ and $x(K) = \mu\epsilon_K + (1-\mu)X$. Thus, by concavity,

$$\mu\epsilon_K + (1-\mu)X = x(K)$$

$$= f(x(K),\ K) \geq \mu f(\epsilon_K, K) + (1-\mu)\ f(X,K)$$

$$> \mu\epsilon_K + (1-\mu)\ f(X,K)$$

$$\Rightarrow X > f(X,K) \quad .$$

Hence, $\chi(K) \neq \emptyset$ implies $\chi(K) = \{x(K)\}$. □

Corollary 2.29. Let $K \in \kappa$ and assume Σ_K holds.

(i) If $\chi(K) = \emptyset$, then $f(X,K) > X$, for all $X > 0$.

(ii) If $\chi(K) \neq \emptyset$, then

$$f(X,K) > X, \qquad X \in (0,\ x(K))$$

$$f(X,K) < X, \qquad X \in (x(K),\ \infty) \quad .$$

Proof. (i) and (ii) follow from Theorem 2.30 and the continuity of $f(\cdot,K)$. □

Theorem 2.30. Let $K \in \kappa$ and assume Σ_K holds.

(i) If $f(K,K) \leq K$, then $X^*(K) = K$.

(ii) If $f(U(K),\ K) \geq U(K)$, then $X^*(K) = U(K)$.

(iii) If $f(K,K) > K$ and $f(U(K),\ K) < U(K)$, then $X^*(K) \in \{[x(K)],\ [x(K)]+1\}$.

Proof. Let $K \in \kappa$.

(i) If $f(K,K) \leq K$, then $\chi(K) \neq \emptyset$ since $f(\epsilon_K,K) > 0$ and $f(\cdot,K)$ is continuous. Hence, by Corollary 2.29, $f(X,K) < X$, $X \in (K,\infty)$. By Theorem 2.20 this implies $\partial C(X,K)/\partial X > 0$, $X \in (K, U(K)]$. Hence $f(K,K) < f(X,K)$, $X \in (K, U(K)]$.

(ii) If $f(U(K), K) \geq U(K)$, then $x(K) \geq U(K)$ by Corollary 2.29. Hence, $f(X,K) > X$, $X \in (K, U(K))$. By Theorem 2.20, this implies $\partial C(X,K)/\partial X < 0$, $X \in (K, U(K))$. Thus, $f(X,K) > f(U(K), K)$, $X \in [K, U(K))$.

(iii) If $f(K,K) > K$ and $f(U(K), K) < U(K)$, then $x(K) \in (K,U(K))$ since $f(\cdot,K)$ is continuous. Let $X' = [x(K)]$. Then, by Corollary 2.29, $f(X,K) > X$, $X \in [K, X')$ and $f(X,K) < X$, $X \in (X'+1, U(K)]$. Hence, by Theorem 2.20, $\partial C(X,K)/\partial X < 0$, $X \in [K, X')$ and $\partial C(X,K)/\partial X > 0$, $X \in (X'+1, U(K)]$. Thus, $C(X,K) > C(X',K)$, $X \in [K,X')$ and $C(X,K) > C(X'+1,K)$, $X \in (X'+1, U(K)]$. Hence, $X^*(K) \in \{X', X'+1\}$. □

The near-monotonicity of $X^*(\cdot)$ is given by

Theorem 2.31. If Σ holds, then $X^*(K) \geq X^*(K')-1$ for all $K,K' \in \kappa$, $K > K'$.

Proof. Let $K,K' \in \kappa$ with $K > K'$. We consider three cases.

Case 1: $f(K',K') \leq K'$. In this case, $X^*(K') = K'$ by Theorem 2.30 and $X^*(K) \geq K > K' = X^*(K')$.

Case 2: $f(U(K'), K') \geq U(K')$. In this case, $X^*(K') = U(K')$ by Theorem 2.30. By Lemma 2.20, $f(U(K'), K) \geq f(U(K'), K') \geq U(K')$. Hence, by Corollary 2.29, $f(X,K) > X$, $X \in (0, U(K'))$ and it then follows from Theorem 2.30 that $X^*(K) \geq U(K') = X^*(K')$.

Case 3: $f(K',K') > K'$ and $f(U(K'), K') < U(K')$. In this case, $X^*(K') \in \{[x(K')], [x(K')]+1\}$ by Theorem 2.30. By Corollary 2.29, $f(X,K') > X$, $X \in (0, x(K'))$. Hence, by Lemma 2.20, $f(X,K) > X$, $X \in (0, x(K'))$. It then follows from Theorem 2.30 that $X^*(K) \geq [x(K')]$, which implies that $X^*(K) \geq X^*(K')-1$. □

With regard to finding $X^*(K)$, there are numerous methods of computing fixed points for a function of one-variable. The technique considered here is function iteration. It will be demonstrated that, under Assumption 2.1', function iteration always determines $X^*(K)$ whenever condition Σ_K holds.

Definition 2.8. For $K \in \kappa$, let $\mathcal{S}(K)$ denote the class of sequences given by

$$X_0 \in [K, U(K)]$$

$$X_{i+1} = \max\{K, \min(f(X_i,K), U(K))\}, \qquad i = 0, 1, 2, \ldots .$$

Lemma 2.32. Let $K \in \kappa$ and $\{X_i\} \in \mathscr{J}(K)$. If g is concave, then $\bar{X} \equiv \lim_{i \to \infty} X_i$ exists and $\bar{X} \in [K, U(K)]$. Furthermore, $\{X_i\}$ is monotone nondecreasing (nonincreasing) if $X_0 \leq f(X_0,K)$ $(X_0 \geq f(X_0,K))$.

Proof. Let $K \in \kappa$ and $\{X_i\} \in \mathscr{J}(K)$. For each i, $K \leq X_i \leq U(K)$, by Definition 2.8; hence, $\{X_i\} \subset [K, U(K)]$. We first show that $\{X_i\}$ is monotone by considering the two cases.

Case 1: $X_0 \leq f(X_0,K)$. In this case, $X_1 = \min\{U(K), f(X_0,K)\} \geq X_0$. If $X_1 = U(K)$, then Lemma 2.22 yields $f(X_1,K) = f(U(K), K) \geq f(X_0,K) \geq U(K) = X_1$, since $X_0 \in [K, U(K)]$. Similarly, if $X_1 = f(X_0,K) \geq X_0$, then $f(X_1,K) \geq f(X_0,K) = X_1$, by Lemma 2.22. In general, assume that $X_i \geq X_{i-1}$ and $f(X_i,K) \geq X_i$. Then $X_{i+1} = \min\{U(K), f(X_i,K)\} \geq X_i$. If $X_{i+1} = U(K)$, then Lemma 2.22 yields $f(X_{i+1},K) = f(U(K), K) \geq f(X_i,K) \geq U(K) = X_{i+1}(K)$, since $X_i \in [K, U(K)]$. Similarly, if $X_{i+1} = f(X_i,K) \geq X_i$, then $f(X_{i+1},K) \geq f(X_i,K) = X_{i+1}$, by Lemma 2.22. Hence, it follows by induction on i that $\{X_i\}$ is nondecreasing whenever $X_0 \leq f(X_0,K)$.

Case 2. $X_0 \geq f(X_0,K)$. In this case, $X_1 = \max\{K, f(X_0,K)\} \leq X_0$. If $X_1 = K$, then Lemma 2.22 yields $f(X_1,K) = f(K,K) \leq f(X_0,K) \leq K = X_1$, since $X_0 \in [K, U(K)]$. Similarly, if $X_1 = f(X_0,K) \leq X_0$, then $f(X_1,K) \leq f(X_0,K) = X_1$, by Lemma 2.22. In general, assume that $X_i \leq X_{i-1}$ and $f(X_i,K) \leq X_i$. Then $X_{i+1} = \max\{K, f(X_i,K)\}$. If $X_{i+1} = K$, then Lemma 2.22 yields $f(X_{i+1},K) = f(K,K) \leq f(X_i,K) \leq K = X_{i+1}$, since $X_i \in [K, U(K)]$. Similarly, if $X_{i+1} = f(X_i,K) \leq X_i$, then

$f(X_{i+1},K) \leq f(X_i,K) = X_{i+1}$, by Lemma 2.22. Hence, it follows by induction on i that $\{X_i\}$ is nonincreasing whenever $X_0 > f(X_0,K)$.

Since $\{X_i\}$ is a monotone sequence and $\{X_i\} \subset [K, U(K)]$, $\bar{X} = \lim_{i \to \infty} X_i$ exists and $\bar{X} \in [K, U(K)]$. □

Theorem 2.33. Let $K \in \kappa$ and $\{X_i\} \in \mathcal{S}(K)$. If Σ_K holds, then $X^*(K) \in \{[\bar{X}], [\bar{X}] + 1\}$ where $\bar{X} \equiv \lim_{i \to \infty} X_i$.

Proof. Let $K \in \kappa$ and $\{X_i\} \in \mathcal{S}(K)$. By the previous lemma, $\bar{X}$ exists and $\bar{X} \in [K, U(K)]$. We consider three possible cases.

Case 1: $f(K,K) \leq K$. By Corollary 2.29, this implies that $f(X,K) < X$, $X \in (K, U(K)]$. Hence, $f(X_0,K) \leq X_0$ and by the previous lemma, $\{X_i\}$ is nonincreasing. Hence, $X_{i+1} = \max\{K, f(X_i,K)\}$, $i = 0, 1, 2, \ldots$. Thus, $\bar{X} = \max\{K, f(\bar{X},K)\}$. If $\bar{X} > K$, then $f(\bar{X},K) < \bar{X}$ and $\max\{K, f(\bar{X},K)\} < \bar{X}$, a contradiction. Therefore, using Theorem 2.30, $\bar{X} = K = X^*(K)$ whenever $f(K,K) \leq K$.

Case 2: $f(U(K), K) \geq U(K)$. By Corollary 2.29, this implies that $f(X,K) > X$, $X \in [K, U(K))$. Hence $f(X_0,K) \geq X_0$ and by the previous lemma, $\{X_i\}$ is nondecreasing. Hence, $X_{i+1} = \min\{U(K), f(X_i,K)\}$, $i = 0, 1, 2, \ldots$. Thus, $\bar{X} = \min\{U(K), f(\bar{X},K)\}$. If $\bar{X} < U(K)$, then $f(\bar{X},K) > \bar{X}$ and $\min\{U(K), f(\bar{X},K)\} > \bar{X}$, a contradiction. Therefore, using Theorem 2.30, $\bar{X} = U(K) = X^*(K)$ whenever $f(U(K),K) \geq U(K)$.

Case 3: $f(K,K) > K$ and $f(U(K), K) < U(K)$. In this case, we consider two subcases to show that $\bar{X} = x(K)$.

Case 3a. $X_0 \leq f(X_0,K)$. In this subcase, $X_{i+1} = \min\{U(K), f(X_i,K)\}$, $i = 0, 1, 2, \ldots$. Thus, $\bar{X} = \min\{U(K), f(\bar{X},K)\}$. But, $f(U(K),K) < U(K)$, so $\bar{X} \neq U(K)$. Hence, $\bar{X} = f(\bar{X},K)$, which implies that $\bar{X} = x(K)$.

Case 3b. $X_0 \geq f(X_0,K)$. In this subcase, $X_{i+1} = \max\{K, f(X_i,K)\}$, $i = 0, 1, 2, \ldots$. Thus, $\bar{X} = \max\{K, f(\bar{X},K)\}$. But $f(K,K) > K$, so $\bar{X} \neq K$. Hence $\bar{X} = f(\bar{X},K)$, which implies that $\bar{X} = x(K)$.

Thus, $\bar{X} = x(K)$ whenever $f(K,K) > K$ and $f(U(K), K) < U(K)$. It follows from Theorem 2.30 that $X^*(K) \in \{[x(K)], [x(K)]+1\} = \{[\bar{X}],[\bar{X}]+1\}$ in this case. □

The following theorem details the convergence properties of the function iteration approach.

Theorem 2.34. Let $K \in \kappa$. Assume that Σ_K holds and that $f(K,K) > K$ and $f(U(K),K) < U(K)$. Let $\{X_i\} \in \mathcal{S}(K)$ with $x(K) > X_0 \geq \max\{\epsilon_K, K\}$. If $(X_{m+2} - X_{m+1}) < (X_{m+1} - X_m)$ for some $m \geq 0$, then $x(K) \in (X_n, X_n+1)$ where:

$$n = m+2 + [(\ln \ell^{-1})^{-1} \ln(U(K)-K) + 1] \tag{2.19}$$

and

$$\ell = (X_{m+2} - X_{m+1})(X_{m+1} - X_m)^{-1} . \tag{2.20}$$

Proof. Since $f(X_0,K) > X_0$ and $x(K) \in (K, U(K))$, $\{X_i\}$ is strictly increasing and $X_{i+1} = f(X_i,K)$, $i \geq 0$. Since $f(\cdot,K)$ is concave over $[X_0,\infty)$, $f(\cdot,K)$ has nonincreasing lefthand derivatives $f'_-(\cdot K)$ over $[X_0,\infty)$ [4]. By concavity, Lemma 2.22, and the hypothesis

$$\begin{aligned} 0 < f'_-(X_{m+1},K) &\leq (f(X_{m+1},K) - f(X_m,K))\,(X_{m+1}-X_m)^{-1} \\ &= (X_{m+2} - X_{m+1})\,(X_{m+1} - X_m)^{-1} \\ &= \ell < 1. \end{aligned}$$

Also, by concavity and Lemma 2.22,

$$\begin{aligned} x(K) = f(x(K),K) &\leq f(X_{m+i},K) + f'_-(X_{m+i},K)\,(x(K) - X_{m+i}) \\ &= X_{m+i+1} + f'_-(X_{m+i},K)\,(x(K) - X_{m+i}) \\ &\leq X_{m+i+1} + \ell(x(K) - X_{m+i}), \qquad i = 1, 2, \ldots \\ \Rightarrow x(K) - X_{m+1+i} &\leq \ell(x(K) - X_{m+i}), \qquad i = 1, 2, \ldots \\ \Rightarrow x(K) - X_{m+1+i} &\leq \ell^{i-1}(x(K) - X_{m+1}), \qquad i = 0, 1, 2, \ldots \\ \Rightarrow x(K) - X_n &\leq \ell^{v}(x(K) - X_{m+1}) \end{aligned}$$

where

$$v \geq (\ln \ell^{-1})^{-1} \ln(U(K)-K)$$

$$= |\ln \ell|^{-1} \ln(U(K)-K)$$

$$= -(\ln \ell)^{-1} \ln(U(K)-K)$$

$$\Rightarrow v \ln \ell \leq -\ln(U(K)-K)$$

$$\Rightarrow \ell^{v} \leq (U(K)-K)^{-1} .$$

Therefore,

$$x(K) - X_n \leq (U(K)-K)^{-1} (x(K) - X_{m+1}) < 1$$

$$\Rightarrow x(K) \in (X_n, X_n+1) ,$$

since $U(K) > x(K) > X_n \geq X_{m+1} > X_0 \geq K$. □

With regard to the condition on (X_m, X_{m+1}, X_{m+2}) in the hypothesis of Theorem 2.34, the following lemma provides an upper bound on m.

Lemma 2.35. Let $K \in \kappa$. Assume that Σ_K holds and that $f(K,K) > K$ and $f(U(K), K) < U(K)$. Let $\{X_i\} \in \mathcal{S}(K)$ with $x(K) > \max\{\epsilon_K, K\}$. Then $(X_{m+2} - X_{m+1}) < (X_{m+1} - X_m)$ for

$$m = [(X_1-X_0)^{-1} (U(K)-K)] + 1 . \tag{2.21}$$

Proof. Since $f(X_0,K) > X_0$ and $x(K) \in (K, U(K))$, $\{X_i\}$ is strictly increasing and $X_{i+1} = f(X_i,K)$, $i \geq 0$. Since $f(\cdot,K)$ is concave over $[X_0,\infty)$, $f(\cdot,K)$ has nonincreasing lefthand derivatives $f'_-(\cdot,K)$ over $[X_0,\infty)$ [4]. We shall show that $f'_-(X_m,K) < 1$; the proof is by contradiction.

Suppose $f'_-(X_m,K) \geq 1$. Then, since $f(\cdot,K)$ is concave and $\{X_i\}$ is increasing, $f'_-(X_i,K) \geq 1$, $i = 0, 1, \ldots, m$. Hence, by concavity,

$$\begin{aligned}
X_{i+1} - X_i &= f(X_i,K) - X_i \\
&\geq f(X_{i-1},K) + f'_-(X_i,K)\,(X_i - X_{i-1}) - X_i \\
&\geq f(X_{i-1},K) + (X_i - X_{i-1}) - X_i \\
&= f(X_{i-1},K) - X_{i-1} \\
&= X_i - X_{i-1}\,, \qquad\qquad i = 1, 2, \ldots, m
\end{aligned}$$

$$\Rightarrow X_{i+1} - X_i \geq X_1 - X_0\,, \qquad\qquad i = 1, 2, \ldots, m$$

$$\Rightarrow X_m - X_0 \geq m(X_1 - X_0)\ .$$

But, by (2.21), the above inequality implies that $X_m \geq X_0 + m(X_1-X_0)$ $\geq X_0 + (U(K)-K) \geq U(K)$. This is a contradiction, since $X_m < \lim_{i\to\infty} X_i$ $= x(K) < U(K)$. Hence, we must conclude that $f'_-(X_m,K) < 1$.

Since $f(\cdot,K)$ is concave and $\{X_i\}$ is increasing,

$$1 > f'_-(X_m,K) \geq (X_{m+1} - X_m)^{-1} (f(X_{m+1},K) - f(X_m,K))$$

$$= (X_{m+1} - X_m)^{-1} (X_{m+2} - X_{m+1}) \quad . \qquad \square$$

Using Theorems 2.33 and 2.34, the Function Iteration Algorithm can be stated as follows.

<u>Algorithm</u> C_4 (Function Iteration, given Σ_K).

(1) If $f(K,K) \leq K$, $X^*(K) = K$ (stop).

(2) If $f(U(K),K) \geq U(K)$, $X^*(K) = U(K)$ (stop).

(3) $X_0 \leftarrow \max\{K, \epsilon_K\}$, $X_1 \leftarrow f(X_0,K)$, $X_2 \leftarrow f(X_1,K)$.

(4) If $(X_2-X_1) < (X_1-X_0)$, go to step (6); otherwise continue.

(5) $X_0 \leftarrow X_1$, $X_1 \leftarrow X_2$, $X_2 \leftarrow f(X_2,K)$; go to step (4).

(6) $\ell \leftarrow (X_2-X_1)(X_1-X_0)^{-1}$, $X \leftarrow X_2$, $i \leftarrow 0$, $i' \leftarrow [|\ln \ell|\ln(U(K)-K) + 1]$.

(7) $X \leftarrow f(X,K)$.

(8) $i \leftarrow i+1$. If $i \leq i'$, go to step (7); otherwise, continue.

(9) $X^*(K) \in \{[X], [X+1], [X+2]\}$ (stop).

(X_0, X_1, X_2) play the roles of (X_m, X_{m+1}, X_{m+2}) as given in Theorem 2.34. By Lemma 2.35, the inequality of step (4) will be satisfied in less than $[(X_1-X_0)^{-1}(U(K)-K) + 1]$ cycles of steps (4) and (5); hence, the algorithm is finite. Step (9) follows from Theorems 2.33 and 2.34; $X^*(K)$ can be determined by computing $C(\cdot,K)$ over the three components.

By Theorem 2.31, additional efficiency can be realized in Algorithm C_4 whenever condition Σ holds and the algorithm is being used recursively to compute $X^*(\cdot)$ over κ. Specifically, for any $K' \in \kappa \sim [K,\infty)$, the initialization of X_0 in step (2) can be replaced

by: $X_0 \leftarrow \max\{K, \epsilon_K, X^*(K')-1\}$ (since $X^*(K')-1 < x(K') \leq x(K)$).

The remainder of this section demonstrates that condition Σ will typically hold for three important classes of concave expansion functions: the concave power functions, the logarithmic functions, and the negative exponential functions.

<u>Example 2.1</u>. Concave Power Functions. This example includes the case of g affine,

$$g(X) = \xi_0 + \xi_1(X+1)^{\xi_2}, \; \xi_0 > 0, \; \xi_1 > 0, \; 0 < \xi_2 \leq 1 ,$$

$$g'(X) = \xi_1 \, \xi_2(X+1)^{\xi_2 - 1} \quad .$$

Denote $a = \ln Z^{-1}$. Then

$$f(X,K) = a^{-1} \, \ln\left\{1 + \frac{\xi_0 + \xi_1(X+1)^{\xi_2} + \phi(K)}{\xi_1 \, \xi_2(X+1)^{\xi_2-1}} \, a\right\} - 1$$

$$\frac{\partial}{\partial X} f(X,K) = \frac{\xi_1^2 \, \xi_2(X+1)^{\xi_2} - (\phi(K)+\xi_0) \, \xi_1\xi_2(\xi_2-1)}{(\xi_1\xi_2)^2 \, (X+1)^{\xi_2} + a(\phi(K) + \xi_0) \, \xi_1\xi_2(X+1) + a\xi_1^2\xi_2(X+1)^{\xi_2+1}} \, .$$

Taking one more derivative yields

$$\operatorname{sgn}\left(\frac{\partial^2}{\partial X^2} f(X,K)\right) = \operatorname{sgn}((\xi_2-1) \, v(X) - a \, \xi_1^4 \, \xi_2^2 \, (X+1)^{2\xi_2}), \qquad (2.22)$$

where

$$v(X) = (\phi(K) + \xi_0)\left\{a\xi_1^3\xi_2^2(1 + (\xi_2+1)\ (X+1)^{\xi_2}) + \xi_1^3\xi_2^4(X+1)^{\xi_2-1}\right.$$

$$\left. + (\phi(K) + \xi_0)\ (\xi_1\xi_2)^2\right\} > 0 \quad .$$

Since $(\xi_2-1) \leq 0$; (2.22) implies $\partial^2/\partial X^2 f(X,K) < 0$ and hence, $f(\cdot,K)$ is concave over $[0,\infty)$ for all $K \in \kappa$. Hence, by Lemma 2.23, Σ_K holds whenever (Σ_K^1) is satisfied and Σ holds whenever (Σ_K^1) is satisfied for all $K \in \kappa$.

(Σ_K^1) will be satisfied whenever

$$f(0,K) = a^{-1}\ \ell n\left(1 + \frac{\xi_0 + \xi_1 + \phi(K)}{\xi_1\ \xi_2}\ a\right) - 1 > 0 \quad .$$

By Lemma 2.23, Σ will hold if

$$b \equiv a^{-1}\ \ell n\left(1 + \frac{\xi_0 + \xi_1}{\xi_1\ \xi_2}\ a\right) - 1 > 0 \ ,$$

since $\phi(0) \geq 0 \Rightarrow f(0,0) \geq b$. Also, by Lemma 2.25, Σ will hold if

$$\frac{g'(0)}{g(0)} = \frac{\xi_1\ \xi_2}{\xi_0 + \xi_1} \leq 1 - \frac{a}{2} \quad .$$

In particular (using Table 2.1), if $\xi_1\xi_2(\xi_0 + \xi_1)^{-1} \leq .1$, then Σ will hold whenever $p \geq .2$ and $r \leq .3$.

Example 2.2. Logarithmic Functions.

$$g(X) = \xi_0 + \xi_1 \ln(X+1) , \qquad \xi_0 > 0,\ \xi_1 > 0$$

$$g'(X) = \xi_1 (X+1)^{-1} .$$

Denote $a = \ln Z^{-1}$. Then

$$f(X,K) = a^{-1} \ln \left(1 + \frac{\xi_0 + \xi_1 \ln(X+1) + \phi(K)}{\xi_1 (X+1)^{-1}} a \right) -1 ,$$

$$\frac{\partial}{\partial X} f(X,K) = \frac{\xi_1 + \xi_0 + \xi_1 \ln(X+1) + \phi(K)}{\xi_1 + a(X+1)\ \{\xi_1 \ln(X+1) + \phi(K)\}} . \qquad (2.23)$$

Taking one more derivative yields

$$\operatorname{sgn}\{\frac{\partial^2}{\partial X^2} f(X,K)\} = \operatorname{sgn}\{v_1(X) - a v_2(X)\}$$

where

$$v_1(X) = \xi_1^2 (X+1)^{-1}$$

and

$$v_2(X) = (\phi(K) + \xi_0)\ \{\xi_1(2 \ln(X+1) + 1) + \phi(K) + \xi_0\}$$
$$+ \xi_1^2 \{1 + \ln(X+1) + (\ln(X+1))^2\} .$$

Since $v_1(\cdot)$ is decreasing and $v_2(\cdot)$ is increasing, $\partial^2/\partial X^2\ f(X,K)$ is decreasing. Furthermore, $v_1(X) \to 0$ and $v_2(X) \to +\infty$ as $X \to \infty$. Hence, either

(a) $v_1(0) - av_2(0) \leq 0$; or

(b) there exists $\epsilon_K > 0$ such that

$$v_1(X) - av_2(X) \begin{cases} > 0, & X \in [0,\epsilon_K) \\ \leq 0, & X \in (\epsilon_K,\infty) \end{cases} .$$

In case (a), $f(\cdot,K)$ is concave over $[0,\infty)$. Since $v_2(0) > \xi_1^2$,

$$v_1(X) > av_2(X) \Rightarrow \xi_1^2 > av_2(X)\ (X+1)$$
$$\geq av_2(0)\ (X+1) > a\xi_1^2(X+1) \quad .$$

Therefore, using (2.23), $v_1(X) > av_2(X) \Rightarrow 1 > a(X+1) \Rightarrow \frac{\partial}{\partial X} f(X,K) > 1$. Thus, either,

(a') $f(\cdot,K)$ is concave over $[0,\infty)$; or

(b') there exists $\epsilon_K > 0$ such that $f(\cdot,K)$ is concave over $[\epsilon_K,\infty)$ and $\frac{\partial}{\partial X} f(\cdot,K) > 1$ over $[0, \epsilon_K)$.

In either case, it follows that Σ_K holds if $f(0,K) > 0$ (and that Σ holds if $f(0,0) > 0$).

Hence, Σ_K will hold if

$$f(0,K) = a^{-1} \ln(1 + \frac{\xi_0 + \phi(K)}{\xi_1} a) - 1 > 0 \quad .$$

By Lemma 2.23, Σ will hold if

$$b \equiv a^{-1} \ln(1 + \frac{\xi_0}{\xi_1} a) - 1 > 0 \quad ,$$

since $\phi(0) \geq 0 \Rightarrow f(0,0) \geq b$. Also, by Lemma 2.25, Σ will hold if

$$\frac{g'(0)}{g(0)} = \frac{\xi_1}{\xi_0} \leq 1 - \frac{a}{2} \quad .$$

In particular (using Table 2.1), if $\xi_1/\xi_0 \leq .1$, then Σ will hold whenever $p \geq .2$ and $r \leq .3$.

<u>Example 2.3</u>. Negative Exponential Functions.

$$g(X) = \xi_0 - \xi_1 e^{-\xi_2(X+1)} \quad , \qquad \xi_0 > \xi_1 > 0, \ \xi_2 > 0$$

$$g'(X) = \xi_2 \xi_1 e^{-\xi_2(X+1)} \quad .$$

Denote $a = \ln Z^{-1}$. Thus

$$f(X,K) = a^{-1} \ln\left(1 + \frac{\xi_0 - \xi_1 e^{-\xi_2(X+1)} + \phi(K)}{\xi_2 \xi_1 e^{-\xi_2(X+1)}} a \right) - 1 ,$$

$$\frac{\partial}{\partial X} f(X,K) = \frac{\xi_2(\xi_0 + \phi(K))}{a(\xi_0 + \phi(K)) + \xi_1(\xi_2 - a) e^{-\xi_2(X+1)}} \quad .$$

Thus, if $\xi_2 \leq a$, then $\partial/\partial X\ f(X,K)$ is nonincreasing and $f(\cdot,K)$ is concave over $[0,\infty)$. Assume $\xi_2 \leq a$.

Then Σ_K will hold whenever

$$f(0,K) = a^{-1} \ln\left(1 + \frac{\xi_0 - \xi_1 e^{-\xi_2} + \phi(K)}{\xi_2 \xi_1 e^{-\xi_2}} a\right) - 1 > 0 \quad .$$

The above inequality will be satisfied if

$$a^{-1} \ln(1 + \frac{\xi_0 + \phi(K)}{\xi_2\xi_1} a) - 1 > 0 \quad .$$

By Lemma 2.23, Σ will hold if

$$b \equiv a^{-1} \ln(1 + \frac{\xi_0}{\xi_2\xi_1} a) - 1 > 0 \quad .$$

Since $\phi(0) \geq 0$ and $e^{-\xi_2} < 1 \Rightarrow f(0,0) > b$. Also, by Lemma 2.25, Σ will hold if

$$\frac{g'(0)}{g(0)} = \frac{\xi_2 \xi_1 e^{-\xi_2}}{\xi_0 + \xi_1 e^{-\xi_2}} \leq 1 - \frac{a}{2} \quad .$$

Since $0 < e^{-\xi_2} < 1$, the above inequality will be satisfied if $\xi_2 \xi_1 \xi_0^{-1} \leq 1 - a/2$. In particular (using Table 4.1), if $\xi_2 \xi_1 \xi_0^{-1} \leq .1$, then Σ will hold whenever $p \geq .2$ and $r \leq .3$.

2.5. Summary

Given the recurrent stochastic capacity expansion models presented previously in [5] and [6], a Policy Improvement algorithm was derived to determine optimal expansion sizes for general expansion functions. For a certain class of expansion functions, which include discrete convex functions, it was shown that each functional $C(\cdot,K)$ was unimodal and that the optimal expansion size $X^*(K) + 1$ increases monotonically with K: an Interval-Bisection algorithm was given to determine $X^*(K)$ for this case. The monotonicity of the optimal expansion size was also demonstrated, prior to integer truncation, for important classes of concave expansion functions which are continuously differentiable; a simple Function Iteration algorithm was derived to determine optimal expansion sizes for this case.

CHAPTER 3

MODEL III: A NON-RECURRENT GENERALIZATION

3.1. Model III

Models I and II were constructed under general assumptions (a) - (d) listed in Section 1.1. In order to illustrate possible generalization for these two models, assumptions (b) - (d) will be simultaneously relaxed for the case of non-modular temporary facilities (i.e., Model I); the generalized model so obtained will be referred to as Model III. Although the generalization will be illustrated using non-modular temporary facilities, similar generalization for the case of modular temporary facilities (i.e., Model II) is also possible; the analogies should be obvious.

It should be emphasized that the material given here is preliminary in nature; its purpose is to indicate the type of extensions that appear to be worthy of more detailed future investigation.

The construction of Model III will proceed under the following basic assumptions:

Assumption 3.1. All costs are time-stationary.

Assumption 3.2. There exists a known integer $\bar{\mathcal{P}} > 0$ which represents the saturation population: an upper bound on the level of system demand.

Assumption 3.3. When the system demand $\mathcal{D}(t) = i$, arrivals occur according to a Poisson process with rate $\lambda_1(i)$ and departures occur (independently) according to a Poisson process with rate $\lambda_2(i)$; $\lambda_1(\bar{\mathcal{D}}) = 0 = \lambda_2(0)$.

Assumption 3.4. Permanent facilities are always used in preference to temporary facilities and temporary facility charges are functions only of the excess demand, the amount of permanent capacity available and the temporary facilities usage limit.

Assumption 3.1 has been implicitly assumed in all previous models. The implication of this assumption is that future advancements in technology will offset inflation (or vice-versa).

Assumption 3.2 differs from all assumptions used in the construction of previous models and can be restated as simply $\mathcal{D}(t) \leq \bar{\mathcal{D}}$, $t \geq 0$. This assumption guarantees that the system will never "explode". Letting $Y(t)$ denote the permanent capacity available at time t, Assumption 3.2 implies that it suffices to bound $Y(t) \leq \bar{\mathcal{D}}$. This assumption was technically unnecessary in previous models, because the assumptions of those models allowed characterizations that were independent of $Y(t)$. The physical interpretation of the parameter $\bar{\mathcal{D}}$ is obvious: $\bar{\mathcal{D}}$ is an upper bound on the ultimate population of the surrounding service area for the system.

Assumption 3.3 generalizes the demand process within the Poisson framework, by allowing the arrival and departure rates to vary according to the system demand level. If the population of the surrounding service area is presumed to be constant, then this assumption can alternatively be viewed as allowing the arrival and departure rates to vary according to the percentage of the surrounding population currently engaging the system's service. However, it is important to note that, given the system demand level, the arrival and departure rates are time-stationary.

The restriction $\lambda_1(\bar{\mathscr{D}}) = 0$ simply states that whenever the system is servicing the entire (saturation) population, then the next event cannot be an arrival. The restriction $\lambda_2(0) = 0$ truncates the demand process; that is, if the system demand is zero, then the next event cannot be a departure.

Denote $\lambda(i) = \lambda_1(i) + \lambda_2(i)$, $p(i) = \lambda_1(i)/\lambda(i)$, and $q(i) = \lambda_2(i)/\lambda(i)$, $0 \leq i \leq \bar{\mathscr{D}}$. Then the demand process can alternatively be characterized by $\{\lambda(i), p(i), q(i)\}$ as follows: if $\mathscr{D}(t) = i$, then the next event occurs according to a Poisson process with rate $\lambda(i)$ with $p(i)$ the probability that the next event is an arrival and $q(i)$ the probability that the next event is a departure [3]. Note that $p(\bar{\mathscr{D}}) = q(0) = 0$ and $p(0) = q(\bar{\mathscr{D}}) = 1$.

Assumption 3.4 states that the temporary facility costs behave according to the non-modular case. The important point here is that these costs are presumed to be independent of the past history of temporary facility utilization. Given the permanent capacity available (Y) and

the temporary facility usage limit (K), the assumption presumes a known unique correspondence between facility costs and the value of the spares level $k(t) = Y(t) - \mathcal{N}(t)$. As demonstrated previously in [5] and [6], this assumption does not usually hold for the case of modular temporary facilities.

Permanent expansion costs will be denoted as $g(X,Y,K)$ which represents the cost of adding $X+1$ units of permanent capacity whenever there are Y units of permanent capacity available prior to the expansion and the temporary facilities usage limit is K prior to the expansion; as before, it will be assumed that $X \geq K$. As in previous models, expansion occurs whenever $\mathcal{N}(t) = Y(t) + K(t)$ (i.e., $k(t) = -K(t)$) and an arrival next occurs. When expansion occurs at time t, the spares level instantaneously increases to $k(t^+) = X - K(t)$ and the permanent capacity available instantaneously increases to $Y(t^+) = Y(t) + X + 1$.

Non-expansion costs will be represented incrementally. $F_k(K,Y)$ will denote the expected incremental (until the next event) discounted costs while the spares level has value k, given that the limit on temporary facilities is currently K and that Y units of permanent capacity are currently available.

Under Assumption 3.2, the following bounds are nonrestrictive for all $t \geq 0$:

$$0 \leq Y(t) + K(t) \leq \bar{\mathcal{N}} \,, \tag{3.1}$$

and

$$-K(t) \leq k(t) \leq Y(t) \,. \tag{3.2}$$

3.2. The Cost Transition Equations

Let $C_k(K,Y)$ denote the minimum expected total discounted costs starting from $k(0) = k$, $K(0) = K$ and $Y(0) = Y$; $-K \leq k \leq Y$, $0 \leq K \leq \bar{\mathfrak{D}}-Y$, $0 \leq Y \leq \bar{\mathfrak{D}}$. For $0 \leq i \leq \bar{\mathfrak{D}}$, denote

$$\alpha(i) = \frac{\lambda(i)\ q(i)}{\lambda(i)+r} \quad \text{and} \quad \beta(i) = \frac{\lambda(i)\ p(i)}{\lambda(i)+r} \quad . \tag{3.3}$$

By conditioning on the time and the type of the next event, transition equations for $\{C_k(K,Y)\}$ can be written in a manner analogous to that used for Model I. Three cases will be considered.

Case 1: $K = Y = 0$. In this case, there is only one transition equation since the only feasible spares level is $k = 0$. Recall that $q(0) = 0$; hence, the first event will be an arrival, triggering an expansion. When expansion occurs, the spares level will increase to X and the permanent capacity will increase to $X+1$. Since $\bar{\mathfrak{D}}$ is an upper bound on the permanent capacity, it suffices to restrict the expansion size by $0 \leq X \leq \bar{\mathfrak{D}}-1$. The costs until the first expansion are given by $F_0(0,0)$. The first expansion cost will be $g(X,0,0)$ and the expected costs thereafter are $C_{X-K}(K', X+1)$ (relative to the expansion time), where K' denotes the temporary facilities usage limit that will be used between the first and second expansion times. By (3.1), it sufficies to restrict $K' + X + 1 \leq \bar{\mathfrak{D}}$, or $0 \leq K' \leq \bar{\mathfrak{D}}-X-1$. Whenever the first expansion time occurs, the costs from that time forward are given by the sum $g(X,0,0) + C_{X-K}(K', X+1)$. We can do no better than minimize this sum over the

feasible values for X and K'. Hence, the single transition equation for this case becomes:

$$C_0(0,0) = F_0(0,0) + \int_0^\infty \min\{g(X,0,0) + C_{X-K}(K', X+1)\}\ p(0)\ \lambda(0)\ e^{-\lambda(0)t}e^{-rt}dt$$

$$= F_0(0,0) + \frac{\lambda(0)\ p(0)}{\lambda(0)+r} \min\{g(X,0,0) + C_{X-K}(K', X+1)\}$$

$$= F_0(0,0) + \beta(0) \min_{\substack{0 \leq X \leq \bar{\mathfrak{D}}-1 \\ 0 \leq K' \leq \bar{\mathfrak{D}}-X-1}} \{g(X,0,0) + C_{X-K}(K', X+1)\}. \quad (3.4)$$

<u>Case 2</u>: $0 < Y+K < \bar{\mathfrak{D}}$. In this case, we consider three subcases.

<u>Case 2a</u>. $k = Y$. In this subcase, the first event must be an arrival, since $q(0) = 0$. Conditioning on the time of the first event yields (recall that $p(0) = 1$):

$$C_Y(K,Y) = F_Y(K,Y) + \int_0^\infty C_{Y-1}(K,Y)\ p(0)\ \lambda(0)\ e^{-\lambda(0)t}\ e^{-rt}\ dt$$

$$= F_Y(K,Y) + \frac{\lambda(0)\ p(0)}{\lambda(0)+r} C_{Y-1}(K,Y)$$

$$= F_Y(K,Y) + \beta(0)\ C_{Y-1}(K,Y)\ . \quad (3.5a)$$

Case 2b: $-K < k < Y$. In this subcase, the first event is an arrival with probability $p(Y-k)$ and a departure with probability $q(Y-k)$. Hence, conditioning on the time and the type of the first event gives

$$
\begin{aligned}
C_k(K,Y) = F_k(K,Y) + \int_0^{\infty} &\{p(Y-k)\ C_{k-1}(K,Y) + q(Y-k)\ C_{k+1}(K,Y)\} \\
&\cdot \lambda(Y-k)\ e^{-\lambda(Y-k)t}\ e^{-rt}\ dt \\
= F_k(K,Y) + \alpha(Y-k)\ &C_{k+1}(K,Y) + \beta(Y-k)\ C_{k-1}(K,Y), \\
&-K < k < Y\ . \qquad (3.5b)
\end{aligned}
$$

Case 2c: $k = -K$. In this subcase, the first event is an arrival with probability $p(Y+K)$ and a departure with probability $q(Y+K)$. If the first event is an arrival, then an expansion of size $X+1$ $(X \geq K)$ is triggered. If expansion occurs, the spares level increases to X and the permanent capacity increases to $Y+X+1$. Proceeding as in Case 1, the minimum expected costs incurred from the expansion time forward are given by

$$
\min_{\substack{K \leq X \leq \bar{\ell}-Y-1 \\ 0 \leq K' \leq \bar{\ell}-Y-X-1}} \{g(X,K,Y) + C_{X-K}(K',\ Y+X+1)\}\ .
$$

Hence, conditioning on the time and type of the first event yields:

$$C_{-K}(K,Y) = F_{-K}(K,Y) + \alpha(K+Y)\, C_{-K+1}(K,Y)$$

$$+ \beta(K+Y) \min_{\substack{K \le X \le \bar{\wp}-Y-1 \\ 0 \le K' \le \bar{\wp}-Y-X-1}} \{g(X,K,Y) + C_{X-K}(K',Y+X+1)\}. \tag{3.5c}$$

The cost transition equations for Case 2 are thus given by (3.5a), (3.5b) and (3.5c).

Case 3: $K+Y = \bar{\wp}$. In this case, the permanent and temporary facility capacities suffice to satisfy the saturation population. Hence, expansion is never necessary in this case. We shall again consider three subcases.

Case 3a: $k = Y$. This subcase is identical to Case 2a. Thus,

$$C_Y(K,Y) = F_Y(K,Y) + \beta(0)\, C_{Y-1}(K,Y) \,. \tag{3.6a}$$

Case 3b: $-K < k < Y$. This subcase is identical to Case 2b. Hence,

$$C_k(K,Y) = F_k(K,Y) + \alpha(Y-k)\, C_{k+1}(K,Y) + \beta(Y-k)\, C_{k-1}(K,Y) \,,$$

$$-K < k < Y \,. \tag{3.6b}$$

Case 3c: $k = -K$. In this subcase, $\wp(0) = Y-k = Y+K = \bar{\wp}$. Hence, since $p(\bar{\wp}) = 0$, the first event will be a departure. Thus, conditioning on the time of the first event yields:

$$C_{-K}(K,Y) = F_{-K}(K,Y) + \alpha(\bar{\mathcal{N}})\ C_{-K+1}(K,Y) \ . \tag{3.6c}$$

The cost transition equations for Case 3 are thus given by (3.6a), (3.6b) and (3.6c).

Using matrix notation, it is possible to write a single unified expression for the cost transition equations (3.4), (3.5) and (3.6).

<u>Definition 3.1</u>. For $0 \leq K \leq \bar{\mathcal{N}}\text{-}Y$ and $0 \leq Y \leq \bar{\mathcal{N}}$, let

$$\mathcal{E}(K,Y) = \begin{cases} \beta(K+Y) \displaystyle\min_{\substack{K \leq X \leq \bar{\mathcal{N}}-Y-1 \\ 0 \leq K' \leq \bar{\mathcal{N}}-Y-X-1}} \{g(X,K,Y) + C_{X-K}(K',Y+X+1)\}, & \text{if } K+Y < \bar{\mathcal{N}} \\ \\ 0\ , & \text{if } K+Y = \bar{\mathcal{N}} \end{cases}$$

Using the above definition, the cost transition equations (3.4) - (3.6) can be rewritten as:

$$C_0(0,0) = F_0(0,0) + \mathcal{E}(0,0)$$

and

$$C_Y(K,Y) - \beta(0)\ C_{Y-1}(K,Y) = F_Y(K,Y)$$

$$-\alpha(Y-k)\ C_{k+1}(K,Y) + C_k(K,Y) - \beta(Y-k)\ C_{k-1}(K,Y) = F_k(K,Y), \quad -K < k < Y$$

$$-\alpha(Y+K)\ C_{-K+1}(K,Y) + C_K(K,Y) = F_{-K}(K,Y) + \mathcal{E}(K,Y)$$

$$0 \leq Y \leq \bar{\mathcal{N}};\ 0 \leq K \leq \bar{\mathcal{N}}\text{-}Y,\ Y+K > 0 \tag{3.7}$$

Definition 3.2. For $m \geq 0$, let $A(m)$ denote the square tridiagonal matrix of dimension $m+1$ with nonzero elements given by

$$A_{ii}(m) = 1 , \qquad i = 0, 1, \ldots, m$$

$$A_{i,i+1}(m) = -\beta(i) , \qquad i = 0, 1, \ldots, m-1$$

$$A_{i,i-1}(m) = -\alpha(i),, \qquad i = 1, 2, \ldots, m .$$

Let $C(K,Y)$ denote the $(K+Y+1)$-vector given by

$$C(K,Y) = (C_Y(K,Y),\ C_{Y-1}(K,Y),\ \ldots,\ C_0(K,Y),\ \ldots,\ C_{-K}(K,Y))^T .$$

Similarly, let $F(K,Y)$ denote the $(K+Y+1)$-vector given by

$$F(K,Y) = (F_Y(K,Y),\ F_{Y-1}(K,Y),\ \ldots,\ F_0(K,Y),\ \ldots,\ F_{-K}(K,Y))^T .$$

Using Definition 3.2 and the above vectors, the cost transition equations (3.7) become

$$A(K+Y)\ C(K,Y) = F(K,Y) + \mathcal{E}(K,Y)\ e(K+Y) ,$$

$$0 \leq K \leq \bar{\mathcal{D}}-Y,\ 0 \leq Y \leq \bar{\mathcal{D}} , \qquad (3.8)$$

where $e(m)$ denotes the $(m+1)$-unit vector with last component unity and all other components zero, $m \geq 0$.

Thus, given $\mathcal{E}(K,Y)$, the costs $C(K,Y)$ can be determined by solving the nonsingular linear system given by (3.8). The next section demonstrates that $\mathcal{E}(K,Y)$ and $C(K,Y)$ can be computed recursively.

3.3. A Backward Recursion for Computing the Minimum Costs

In this section, a backward recursion is derived to solve (3.8) for all feasible values of K and Y.

Definition 3.3. For $0 \le Y \le \bar{\mathcal{D}}$, define

$$Z_k(Y) = \min_{0 \le K' \le \bar{\mathcal{D}}-Y} C_k(K',Y), \qquad 0 \le k \le Y \ . \tag{3.9}$$

Combining Definitions 3.1 and 3.3 gives

$$\begin{aligned}
\mathcal{E}(K,Y) &= \beta(K+Y) \min_{\substack{K \le X \le \bar{\mathcal{D}}-Y-1 \\ 0 \le K' \le \bar{\mathcal{D}}-Y-X-1}} \{g(X,K,Y) + C_{X-K}(K',Y+X+1)\} \\
&= \beta(K+Y) \min_{K \le X \le \bar{\mathcal{D}}-Y-1} \{g(X,K,Y) + \min_{0 \le K' \le \bar{\mathcal{D}}-Y-X-1} C_{X-K}(K',Y+X+1)\} \\
&= \beta(K+Y) \min_{K \le X \le \bar{\mathcal{D}}-Y-1} \{g(X,K,Y) + Z_{X-K}(Y+X+1)\} , \\
&\qquad 0 \le K < \bar{\mathcal{D}}-Y,\ 0 \le Y < \bar{\mathcal{D}} \ .
\end{aligned} \tag{3.10}$$

The recursion begins with the computation of $C(0, \bar{\mathcal{D}})$. By Definition 3.1, $\mathcal{E}(0, \bar{\mathcal{D}}) = 0$. Hence, from (3.8),

$$C(0, \bar{\mathfrak{D}}) = A^{-1}(\bar{\mathfrak{D}})\ F(0, \bar{\mathfrak{D}}) \quad . \tag{3.11}$$

By (3.1), $Y = \bar{\mathfrak{D}}$ implies $K = 0$. Hence, (3.9) reduces to

$$\mathcal{Z}_k(\bar{\mathfrak{D}}) = C_k(0, \bar{\mathfrak{D}}) = A^{-1}_{k.}(\bar{\mathfrak{D}})\ F(0, \bar{\mathfrak{D}}) \ , \qquad 0 \le k \le \bar{\mathfrak{D}} \quad . \tag{3.12}$$

In order to illustrate a typical iteration of the recursion, let $Y < \bar{\mathfrak{D}}$ be arbitrary and suppose that $\{\mathcal{Z}_k(Y'),\ 0 \le k \le Y'\}$ are known for $Y+1 \le Y' \le \bar{\mathfrak{D}}$. Then, for each $K = 0, 1, \ldots, \bar{\mathfrak{D}}-Y$, $\mathcal{E}(K,Y)$ can be found using (3.10) and $C(K,Y)$ can be computed as

$$C(K,Y) = A^{-1}(K+Y)\ \{F(K,Y) + \mathcal{E}(K,Y)\ e(K+Y)\} \quad . \tag{3.13}$$

$\{\mathcal{Z}_k(Y),\ 0 \le k \le Y\}$ can then be calculated using (3.9). At this point, the quantities $\{\mathcal{Z}_k(Y'),\ 0 \le k \le Y'\}$ are now known for $Y \le Y' \le \bar{\mathfrak{D}}$ and the next iteration of the recursion begins for permanent capacity level $Y-1$.

The recursion continues for $Y = \bar{\mathfrak{D}}-1, \bar{\mathfrak{D}}-2, \ldots, Y_0+1$, where Y_0 denotes the initial permanent capacity of the system. Denoting by K_0 the initial temporary facilities usage limit, $\mathcal{E}(K_0, Y_0)$ can then be computed using (3.10). Finally, denoting by k_0 the initial spares level, the desired minimum cost is given by

$$C_{k_0}(K_0, Y_0) = A^{-1}_{k_0.}(K_0+Y_0)\ \{F(K_0,Y_0) + \mathcal{E}(K_0,Y_0)\ e(K_0+Y_0)\} \quad .$$

The sequence of optimal actions resulting in the above minimum cost can then be determined by a forward lookup procedure: find the sequence of expansion sizes, permanent capacity levels and temporary facilities usage limits which are the minimands in (3.9) and (3.10) leading to $C_{k_0}(K_0,Y_0)$. This is a standard procedure in dynamic programming which is normally accomplished by storing all minimands along with a set of pointers during the backward recursions [2].

Recall that the system of equations (3.8) is tridiagonal. Given the special structure of a tridiagonal system, one would normally not compute the entire inverse matrix to solve the system (see, for instance, [5]). However, in the case at hand, each inverse matrix is used repeatedly; in particular, $A^{-1}(m)$ is used $(m - Y_0 + 1)$ times for each $m \in \{Y_0+1, \ldots, \bar{\mathscr{Y}}\}$. Hence, it may be advantageous in the general case to compute the necessary inverse matrices (at least in part) once and for all, prior to initiating the recursion. The following theorem provides a separate recursion for doing so.

Theorem 3.1. $A^{-1}(0) = [1]$ and for $m \geq 0$,

$$A^{-1}(m+1) = \begin{bmatrix} \alpha(m) & \beta(m) \\ \gamma(m) & \delta(m) \end{bmatrix}, \tag{3.14}$$

where

$$\delta(m) = (1 - \alpha(m+1)\ \beta(m)\ A^{-1}_{mm}(m))^{-1} \in \mathbb{R}\ , \tag{3.14a}$$

$$\mathcal{C}(m) = \delta(m)\ \alpha(m+1)\ A^{-1}_{m.}(m); \quad \mathcal{C}^{T}(m) \in \mathbb{R}^{m+1}\ , \tag{3.14b}$$

$$\mathcal{B}(m) = \delta(m)\ \beta(m)\ A^{-1}_{.m}(m) \in \mathbb{R}^{m+1}\ , \tag{3.14c}$$

$$\mathcal{A}(m) = A^{-1}(m) + \delta(m)\ \alpha(m+1)\ \beta(m)\ A^{-1}_{.m}(m)\ A^{-1}_{m.}(m) \in \mathbb{R}^{(m+1)\times(m+1)}\ . \tag{3.14d}$$

<u>Proof</u>: By Definition 3.1,

$$A(m+1) = \begin{bmatrix} A(m) & B(m) \\ C(m) & 1 \end{bmatrix}\ , \tag{3.15}$$

where

$$B(m) = (0,\ 0,\ \ldots,\ 0,\ -\beta(m))^{T} \in \mathbb{R}^{m+1}\ , \tag{3.15a}$$

$$C(m) = (0,\ 0,\ \ldots,\ 0,\ -\alpha(m+1));\ C^{T}(m) \in \mathbb{R}^{m+1}\ . \tag{3.15b}$$

Also, $A(m)$ is nonsingular and so $A^{-1}(m)$ exists. $A(m+1)$ is nonsingular, so $A^{-1}(m+1)$ exists; suppose that $A^{-1}(m+1)$ is partitioned as in (3.14). Then, since $A(m+1)\ A^{-1}(m+1) = I$, it follows from (3.14) and (3.15) that

$$A(m)\ \mathcal{A}(m) + B(m)\ \mathcal{C}(m) = I \quad , \tag{3.16}$$

$$A(m)\ \mathcal{B}(m) + B(m)\ \delta(m) = (0, \ldots, 0)^T \ , \tag{3.17}$$

$$C(m)\ \mathcal{A}(m) + \mathcal{C}(m) = (0, \ldots, 0) \ , \tag{3.18}$$

$$C(m)\ \mathcal{B}(m) + \delta(m) = 1 \quad . \tag{3.19}$$

From (3.17),

$$A(m)\ \mathcal{B}(m) = -\delta(m)\ B(m) \Rightarrow \mathcal{B}(m) = -\delta(m)\ A^{-1}(m)\ B(m) \quad . \tag{3.20}$$

Substituting (3.20) into (3.19) gives

$$-\delta(m)\ C(m)\ A^{-1}(m)\ B(m) + \delta(m) = 1$$

$$\Rightarrow \delta(m) = (1 - C(m)\ A^{-1}(m)\ B(m))^{-1} \quad . \tag{3.21}$$

Using (3.15a),

$$A^{-1}(m)\ B(m) = -\beta(m)\ A^{-1}_{.m}(m) \quad . \tag{3.22}$$

Hence, using (3.15b),

$$C(m)\ A^{-1}(m)\ B(m) = \alpha(m+1)\ \beta(m)\ A^{-1}_{mm}(m) \quad .$$

Substituting the above expression into (3.21) gives

$$\delta(m) = (1 - \alpha(m+1)\ \beta(m)\ A^{-1}_{mm}(m))^{-1} \quad ,$$

which verifies (3.14a). Using (3.16),

$$A(m)\ \mathfrak{a}(m) = I - B(m)\ \mathfrak{c}(m) \Rightarrow \mathfrak{a}(m) = A^{-1}(m) - A^{-1}(m)\ B(m)\ \mathfrak{c}(m). \tag{3.23}$$

From (3.18), (3.21) and (3.23),

$$\begin{aligned} & C(m)\ \mathfrak{a}(m) + \mathfrak{c}(m) \\ & \quad = C(m)\ (A^{-1}(m) - A^{-1}(m)\ B(m)\ \mathfrak{c}(m)) + \mathfrak{c}(m) \\ & \quad = C(m)\ A^{-1}(m) - (C(m)\ A^{-1}(m)\ B(m) - 1)\ \mathfrak{c}(m) = 0 \\ & \Rightarrow C(m)\ A^{-1}(m) = (C(m)\ A^{-1}(m)\ B(m) - 1)\ \mathfrak{c}(m) \\ & \Rightarrow \mathfrak{c}(m) = (C(m)\ A^{-1}(m)\ B(m) - 1)^{-1}\ C(m)\ A^{-1}(m) \\ & \quad = -\delta(m)\ C(m)\ A^{-1}(m) \ . \end{aligned} \tag{3.24}$$

Using (3.15b),

$$C(m)\ A^{-1}(m) = -\alpha(m+1)\ A^{-1}_{m.}(m) \ . \tag{3.25}$$

Substituting (3.25) into (3.24) yields

$$\mathcal{C}(m) = \delta(m)\ \alpha(m+1)\ A^{-1}_{m\cdot}(m) \ , \tag{3.26}$$

which verifies (3.14b). Substituting (3.22) into (3.20) gives

$$\mathcal{B}(m) = \delta(m)\ \beta(m)\ A^{-1}_{\cdot m}(m) \ ,$$

which verifies (3.14c). Using (3.22) and (3.26) yields

$$\begin{aligned} A^{-1}(m)\ B(m)\ \mathcal{C}(m) &= -\beta(m)\ A^{-1}_{\cdot m}(m)\ \delta(m)\ \alpha(m+1)\ A^{-1}_{m\cdot}(m) \\ &= -\delta(m)\ \alpha(m+1)\ \beta(m)\ A^{-1}_{\cdot m}(m)\ A^{-1}_{m\cdot}(m) \quad . \end{aligned}$$

Substituting the above expression into (3.23) yields

$$\mathcal{A}(m) = A^{-1}(m) + \delta(m)\ \alpha(m+1)\ \beta(m)\ A^{-1}_{\cdot m}(m)\ A^{-1}_{m\cdot}(m) \quad ,$$

where verifies (3.14d) and completes the proof. □

In the general case being treated here, the minimum cost vectors $C(K,Y)$ must be computed for <u>all</u> feasible pairs (K,Y) in order to perform the minimization (3.9). This necessitates a great deal of computation for even moderate size valves of $\bar{\mathfrak{D}}$, as well as substantial amounts of intermediate storage. However, in many actual applications, it appears that the computation and storage requirements can be significantly reduced.

Specifically, if the incremental expected operating costs for permanent facilities are independent of the parameter K, then significant computational economies accrue. This is demonstrated in the next section.

3.4. A Simplified Backward Recursion for Semi-Constant Expected Incremental Operating Costs

In this section, it is demonstrated that significant computational economies arise whenever the expected incremental operating costs are at least partially independent of the parameter K.

Definition 3.4. The expected incremental operating costs will be said to be semi-constant in K if $F_k(K,Y) = F_k(Y)$, $0 < k \leq Y$, $0 \leq K \leq \bar{\mathcal{D}}-Y$, $0 \leq Y \leq \bar{\mathcal{D}}$. The expected incremental operating costs will be said to be constant in K if $F_k(K,Y) = F_k(Y)$, $-K \leq k \leq Y$, $0 \leq K \leq \bar{\mathcal{D}}-Y$, $0 \leq Y \leq \bar{\mathcal{D}}$.

Obviously, if the expected incremental operating costs are constant in K, then they are also semi-constant in K. Recall that whenever $k \geq 0$, the permanent capacity sufficies to serve demand. Hence, for $k > 0$, $F_k(K,Y)$ will represent the expected incremental operating costs for serving $Y-k$ customers using permanent facilities; it is difficult to imagine practical situations where this expected cost might depend on the temporary facilities usage limit K. Therefore, it appears that the assumption of expected incremental operating costs which are semi-constant in K is extremely viable in practice.

With regard to the presumption of expected incremental operating costs that are constant in K, first note that this presumption does <u>not</u> rule out the possibility of extraordinary disposal costs for the temporary facilities just prior to a permanent expansion. Since the expansion cost function g is parameterized in K, these extraordinary costs can be accounted for there. It is, however, possible to imagine situations where an increase in K will place additional stresses on the entire temporary portion of the system; in such cases, it is hypothesized that the expected incremental operating costs $F_k(K,Y)$ might increase with K for $k \leq 0$.

<u>Theorem 3.2</u>. If the expected incremental costs are semi-constant in K, then $C_k(K,Y) \leq C_k(K',Y)$ if and only if $C_j(K,Y) \leq C_j(K',Y)$, $0 \leq j, k \leq Y$, $0 \leq K, K' \leq \bar{\mathcal{N}}-Y$, $0 \leq Y < \bar{\mathcal{N}}$.

<u>Proof</u>: Let $0 \leq Y < \bar{\mathcal{N}}$, $0 \leq K, K' \leq \bar{\mathcal{N}}-Y$ and $0 \leq j, k \leq Y$ be arbitrarily chosen. To avoid trivialities, assume that $K \neq K'$ and $j \neq k$. Without loss of generality, assume $k > j$.

Let $T_k(j, Y)$ denote the first time that the spares level reaches j, given that the initial spares level is k and that the initial permanent capacity level is Y. The distribution of $T_k(j,Y)$ is a function of the demand process parameters $\{\lambda(i), p(i), q(i), i = 0, 1, \ldots, Y-j-1\}$ and is therefore independent of the values K and K'.

Let $G_k(j,Y)$ denote the expected discounted costs over the time interval $[0, T_k(j,Y)]$. Since $j \geq 0$, no expansion costs are incurred

during this time interval. Hence, $G_k(j,Y)$ is a function of the demand process parameters $\{\lambda(i),\ p(i),\ q(i),\ i = 0,\ 1,\ \ldots,\ Y-j-1\}$ and the expected incremental costs $\{F_\ell(Y),\ \ell = j+1,\ \ldots,\ Y\}$. Thus, $G_k(j,Y)$ is independent of the values K and K'.

From the above definitions,

$$C_k(K,Y) = G_k(j,Y) + C_j(K,Y)\ E[e^{-rT_k(j,Y)}] \ ,$$

and

$$C_k(K',Y) = G_k(j,Y) + C_j(K',Y)\ E[e^{-rT_k(j,Y)}] \ .$$

Since $k \neq j$, $E[e^{-rT_k(j,Y)}] > 0$. Hence, $C_k(K,Y) \leq C_k(K',Y)$ if and only if $C_j(K,Y) \leq C_j(K',Y)$. □

<u>Corollary 3.3</u>. If the expected incremental costs are semi-constant in K, then there exists $K^*(\cdot)$ such that

$$Z_k(Y) = C_k(K^*(Y),\ Y) \ , \qquad 0 \leq k \leq Y$$

for $0 \leq Y \leq \bar{\mathcal{N}}$.

<u>Proof</u>: For $Y = \bar{\mathcal{N}}$, $K^*(\bar{\mathcal{N}}) = 0$ by (5.12). For $0 \leq Y < \bar{\mathcal{N}}$, define $K^*(Y)$ by

$$C_Y(K^*(Y),\ Y) = \min_{0 \leq K \leq \bar{\mathcal{N}}-Y} C_Y(K,Y) \ .$$

It then follows from Theorem 3.2 that

$$C_k(K^*(Y),\ Y) = \min_{0 \le K \le \bar{\mathfrak{D}}-Y} C_k(K,Y) = \mathcal{Z}_k(Y), \quad 0 \le k \le Y. \quad \square$$

The physical interpretation of Corollary 3.3 is that there exists a temporary facility usage limit $K^*(Y)$ that is optimal for each Y, independent of the starting spares level $k \ge 0$.

From a computational point-of-view, the above corollary is very significant. The corollary states that it is unnecessary to compute the entire set of cost vectors $\{C(K,Y),\ 0 \le K \le \bar{\mathfrak{D}}-Y\}$ in order to perform the minimization (3.9) determining $\{\mathcal{Z}_k(Y),\ 0 \le k \le Y\}$. Rather, it suffices to compute only the single elements $\{C_j(K,Y),\ 0 \le K \le \bar{\mathfrak{D}}-Y\}$ for some $j \ge 0$ in order to determine $\{\mathcal{Z}_k(Y),\ 0 \le k \le Y\}$. In that which follows, we shall (arbitrarily) assume that $j = Y$.

By (3.8),

$$C_Y(K,Y) = A_{0.}^{-1}(K+Y)\ \{F(K,Y) + \mathcal{E}(K,Y)\ e(K+Y)\}\ . \tag{3.27}$$

By Theorem 3.1, $A_{0.}^{-1}(0) = [1]$ and for $m \ge 0$,

$$A_{0.}^{-1}(m+1) = [\alpha_{0.}(m),\ \beta_0(m)]\ , \tag{3.28}$$

where

$$\mathcal{A}_{0.}(m) = A^{-1}_{0.}(m) + \delta(m)\ \alpha(m+1)\ \beta(m)\ A^{-1}_{0m}(m)\ A^{-1}_{m.}(m) \quad , \tag{3.28a}$$

$$\mathcal{B}_{0}(m) = \delta(m)\ \beta(m)\ A^{-1}_{0m}(m)\ . \tag{3.28b}$$

Hence, in order to recursively compute $\{C_Y(K,Y),\ 0 \leq K \leq \bar{\mathcal{D}}\text{-}Y\}$ it suffices to recursively calculate $\{A^{-1}_{0.}(K+Y),\ A^{-1}_{K+Y,.}(K+Y),\ 0 \leq K \leq \bar{\mathcal{D}}\text{-}Y\}$. By Theorem 3.1, $A^{-1}_{0.}(0) = [1]$ and for $m \geq 0$,

$$A^{-1}_{m+1,.}(m+1) = [\mathcal{C}(m),\ \delta(m)]\ , \tag{3.29}$$

where

$$\delta(m) = (1 - \alpha(m+1)\ \beta(m)\ A^{-1}_{mm}(m))^{-1}\ , \tag{3.29a}$$

$$\mathcal{C}(m) = \delta(m)\ \alpha(m+1)\ A^{-1}_{m.}(m)\ . \tag{3.29b}$$

Combining (3.28) and (3.29), the first and last rows of $A^{-1}(\cdot)$ can be computed recursively as $(A^{-1}_{0.}(0) = [1])$:

$$\begin{aligned}
\delta(m) &= (1 - \alpha(m+1)\ \beta(m)\ A^{-1}_{mm}(m))^{-1}\ , \\
\mathcal{C}(m) &= \delta(m)\ \alpha(m+1)\ A^{-1}_{m.}(m), \\
A^{-1}_{0.}(m+1) &= [A^{-1}_{0.}(m),\ 0] + A^{-1}_{0m}(m)\ \beta(m)\ [\mathcal{C}(m),\ \delta(m)]\ , \\
A^{-1}_{m+1,.}(m+1) &= [\mathcal{C}(m),\ \delta(m)]\ .
\end{aligned} \tag{3.30}$$

In order to describe a typical iteration of the backward algorithm, when the expected incremental costs are semi-constant in K, assume that the first rows of $A^{-1}(\cdot)$ are pre-calculated using (3.30) prior to initiating the algorithm.

As before, the recursion begins by solving the system

$$C(0, \bar{\mathcal{S}}) = A^{-1}(\bar{\mathcal{S}})\ F(0, \bar{\mathcal{S}}) \tag{3.31}$$

for $C_k(0, \bar{\mathcal{S}})$, $0 \leq k \leq \bar{\mathcal{S}}$. Since $K = 0$ is the only possibility when $Y = \bar{\mathcal{S}}$, $K^*(\bar{\mathcal{S}}) = 0$

Let $Y < \bar{\mathcal{S}}$ be arbitrary and assume that $\{K^*(Y),\ Y < Y' \leq \bar{\mathcal{S}}\}$ are known. By Corollary 3.3, $Z_k(Y') = C_k(K^*(Y'), Y')$, $0 \leq k \leq Y'$, $Y < Y' \leq \bar{\mathcal{S}}$. Then the iteration for Y proceeds as follows:

(1) $K \leftarrow 0$, $C_Y^* \leftarrow \infty$.

(2) $\mathcal{E}(K,Y) \leftarrow \beta(K+Y) \min\limits_{K \leq X \leq \bar{\mathcal{S}}-Y-1} \{g(X,K,Y) + Z_{X-K}(Y+X+1)\}$.

(3) $C_Y(K,Y) \leftarrow A_{0.}^{-1}(K+Y)\ \{F(K,Y) + \mathcal{E}(K,Y)\ e(K+Y)\}$.

(4) If $C_Y(K,Y) \geq C_Y^*$, go to step (6); otherwise continue.

(5) $C_Y^* \leftarrow C_Y(K,Y)$, $K^*(Y) \leftarrow K$.

(6) $K \leftarrow K+1$. If $K \leq \bar{\mathcal{S}}-Y$, go to step (2); otherwise continue.

(7) $Z_k(Y) = C_k(K^*(Y), Y)$, $0 \leq k \leq Y$.

In this simplified recursion, the partial cost vectors $(C_Y(K^*(Y), Y), \ldots, C_0(K^*(Y), Y))^T$ must be computed for each Y (for $Y = \bar{\mathcal{S}}$ to initiate the recursion and in step (7) of each iteration for $Y < \bar{\mathcal{S}}$).

This computation could be performed by using a specialized back-substitution scheme for tridiagonal systems similar to that used for Model I.

3.5. Summary

It has been shown here that many of the assumptions for the simple recurrent models of earlier papers can be relaxed. For a given saturation population, the revised Model III can be solved by backward recursion to yield an optimal sequence of expansion sizes and temporary facility usage limits. The price paid for this generalization is that of greater data requirements (arrival and departure rate estimates are necessary for each demand level), increased computation, and the absence of qualitative results. As shown in the previous section, the backward recursion appears to be computationally feasible under reasonable assumptions concerning the permanent facility operating costs. It is conjectured that the consideration of specific forms for costs and demand rates in the revised model will prove to be a fruitful path of future inquiry.

REFERENCES

[1] Cyrus Derman, *Finite State Markovian Decision Processes*, Academic Press, New York, 1970.

[2] Frederick S. Hillier and Gerald J. Lieberman, *Introduction to Operations Research*, Holden-Day, San Francisco, 1968.

[3] Sheldon M. Ross, *Applied Probability Models with Optimization Applications*, Holden-Day, San Francisco, 1970.

[4] H. L. Royden, *Real Analysis*, MacMillan, Toronto, 1968.

[5] R. Scott Shipley, "A Stochastic Capacity Expansion Model: Non-Modular Temporary Facilities", Stanford University, Department of Operations Research, Technical Report No. 178, September 1976.

[6] R. Scott Shipley, "A Stochastic Capacity Expansion Model: Modular Temporary Facilities", Stanford University, Department of Operations Research, Technical Report No. 179, October 1976.

A033428
343

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 180	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) (6) Optimization of Recurrent Stochastic Capacity Expansion Models and Generalization to a Non-Recurrent Model.		5. TYPE OF REPORT & PERIOD COVERED (9) Technical Report,
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) (10) R. Scott Shipley		8. CONTRACT OR GRANT NUMBER(s) (15) N00014-75-C-0561
9. PERFORMING ORGANIZATION NAME AND ADDRESS Dept. of Operations Research and Dept. of Statistics, Stanford University, Stanford, California 94305		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS (NR-042-002)
11. CONTROLLING OFFICE NAME AND ADDRESS Statistics and Probability Program Code 436 Office of Naval Research Arlington, Virginia 22217		12. REPORT DATE October 11, 1976
		13. NUMBER OF PAGES 80
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) (11) 11 Oct 76 (12) 85p.		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION IS UNLIMITED.

(14) TR-180

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

BACKLOGGING, CAPACITY EXPANSION, INVENTORY THEORY, JOBLETTING, OVERLOADING, POISSON PROCESSES

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

See reverse side.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

332580

1B

20. Abstract

for

In two previous papers, recurrent capacity expansion strategies of the form (X,K) are considered, where X+1 represents the expansion size undertaken whenever excess demand reaches the value K+1. The determination of optimal expansion sizes X*(K)+1 is considered here, for fixed values K. A Policy Improvement algorithm is derived to determine optimal expansion sizes for general expansion functions. For a certain class of expansion functions, which includes discrete convex functions, it is shown that the expected discounted costs are unimodal in the expansion size and that the optimal expansion size X*(K)+1 increases monotonically with K; an Interval-Bisection algorithm is given to determine X*(K) for this case. The monotonicity of the optimal expansion size is also demonstrated, prior to integer-truncation, for important classes of concave expansion functions which are continuiously differentiable; a simple Function Iteration algorithm is derived to determine optimal expansion sizes for this case.

It is then shown that many of the assumptions for the simple recurrent models of earlier papers can be relaxed. For a given upper limit on the ultimate demand level, a revised model is introduced that can be solved by backward recursion to yield an optimal sequence of strategies, each of form (X,K). It is shown that the backward recursion is computationally feasible under reasonable assumptions concerning facility operating costs.